

Memorandum 88330

1N-03

157148

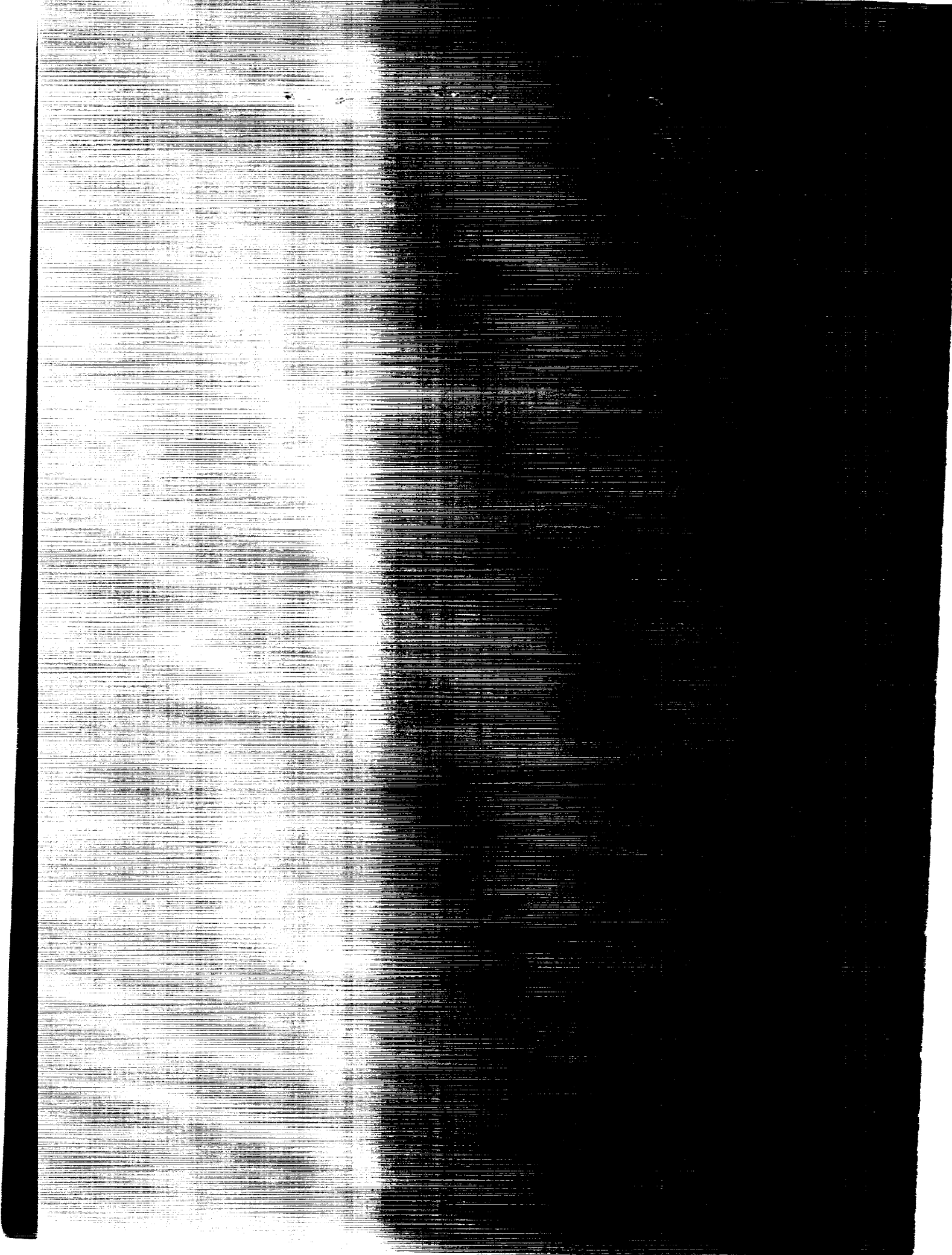
Simulation Human Factors Research Operations

Hennessy,
Yreuls,

(NASA-TM-88330) USING FULL-MISSION
SIMULATION FOR HUMAN FACTORS RESEARCH IN AIR
TRANSPORT OPERATIONS (NASA) 85 P CSCL GSH

N88-10037

G3/03 Unclass
0157148



Using Full-Mission Simulation for Human Factors Research in Air Transport Operations

Harry W. Orlady
Orlady Associates, Inc.
Los Gatos, California

Robert T. Hennessy, Richard Obermayer,
and Donald Vreuls
Vreuls Research Corporation
Thousand Oaks, California

Miles R. Murphy
Ames Research Center
Moffett Field, California



National Aeronautics
and Space Administration

Scientific and Technical
Information Division

TABLE OF CONTENTS

Chapter	Page
1 INTRODUCTION	1
PURPOSE	1
PROBLEM STATEMENT	1
PROCEDURAL APPROACH	1
LITERATURE REVIEW	2
FIELD INTERVIEWS	2
REPORT SCOPE AND ORGANIZATION	5
2 PLANNING AND CONDUCT OF SIMULATION-BASED RESEARCH	5
INTRODUCTION	5
PROBLEM DEFINITION	5
Establishing the Operational Need	5
Establishing the Research Objective	6
STUDY PLAN	6
Framing the General Approach	6
Describing the Operational Conditions	7
Defining the Variables and Developing the Preliminary Scenario	7
Performing the Sensitivity Analysis	8
EXPERIMENTAL STRATEGY	8
Experimental Design	8
Performance Measurement	9
TEST PREPARATION	9
Simulator Capabilities	9
Conflicts Between Simulator Capabilities and Research Requirements	9
Coordination with Simulator Support Personnel	10
Scenario Development	10
Procedures	11
Data Collection System	11
Subject Recruitment	12
Subject Training	12
Pretesting	12
DATA COLLECTION	13
DATA ANALYSIS	13
INTERPRETATION OF DATA	13
CONCLUSION	15
3 GUIDELINES FOR LINE-ORIENTED SIMULATION RESEARCH	15
INTRODUCTION	15
FOUNDATIONS OF LOS RESEARCH	15
Commonality of LOS for Training and for Research	16
Differences Between LOS for Training and LOS for Research	16
WHEN TO USE LOS	16
ADVANTAGES OF LINE-ORIENTED FMS RESEARCH	17
COMPOSITION OF THE RESEARCH TEAM	17
DEVELOPING THE SCENARIO	18
RUNNING THE EXPERIMENT	19
RESEARCH SUBJECTS	19
Subject Selection	20
Subject Training	20

3 CONTINUED

THE SCENARIO	21
A Basic Limitation	21
Elements of Successful Research Scenarios	21
The Operational Problem	21
Real-World Performance	22
A Final Comment on the Importance of Scenario Realism	22
THE SCENARIO SCRIPT	23
SPECIFIC CONSIDERATIONS	23
Route Selection and Scenario Reality	24
Navigational Aids and Communications Services	24
ATC Communications	24
Flight Planning, Dispatch, and Preflight	24
Interaction with Cabin Crew and Passengers	25
Pacing, Tempo, and Quiet Periods	25
TRAINING OF THE OPERATING TEAM AND SCENARIO TESTING	25
PREFLIGHT BRIEFING FOR PARTICIPATING CREWS	25
RUNNING THE SCENARIO	26
Deviations from the Scenario	26
Unexpected Poor Performance	27
Simulator Crashes	27
Simulator Problems	27
DEBRIEFING	28
LEAD-TIME CONSIDERATIONS	28
PREIMPLEMENTATION CHECKLIST FOR LOS RESEARCH	29
Conceptual Stage	29
Development Stage	29
A FINAL COMMENT	29
4 RESEARCH RECOMMENDATIONS	30
INTRODUCTION	31
FULL-MISSION RESEARCH VALIDITY TESTING	31
Problem	31
Approach	31
Expected Results	31
Alternative Approaches	32
ALTERNATIVE FORMS OF VALID SIMULATION	32
Problem	32
Approach	32
Expected Results	33
OPTIMIZATION OF FULL-MISSION RESEARCH	34
Problem	34
Approach	34
Expected Results	34
HUMAN PERFORMANCE MODEL DEVELOPMENT	34
Problem	34
Approach	34
Expected Results	35
SUBJECTIVE MEASURES OF FIDELITY	35
Problem	35
Approach	35
Expected Results	35
INTEGRATION OF RESEARCH EFFORTS	35
CONCLUSIONS	35
	36

Appendix

A	FIELD INTERVIEW – NASA STUDY OF PREFLIGHT AND POSTFLIGHT OPERATIONS	37
	IN SHORT-HAUL OPERATIONS – <i>Dr. H. Clayton Foushee</i>	37
	RESEARCH GOAL	37
	GENERAL DESCRIPTION OF SCENARIO	37
	SPECIFIC SCENARIO CONSIDERATIONS	37
	Subjects	38
	Observer/Ancillary Support Personnel	38
	Flight Planning	38
	Weather	38
	Simulator Equipment	38
	High-Low-Workload Periods	38
	Malfunctions	39
	Ethics	39
	Debriefing	39
	SCENARIO GENERATION PROCEDURES	
B	FIELD INTERVIEW – NASA FULL-MISSION SIMULATION STUDY OF CREW COORDINATION	41
	AND DECISION MAKING – <i>Miles R. Murphy</i>	41
	RESEARCH GOAL	41
	GENERAL DESCRIPTION OF SCENARIO	41
	SPECIFIC SCENARIO CONSIDERATIONS	41
	Simulator	41
	Subjects	42
	Experimenter Team	42
C	APPLIED BEHAVIORAL RESEARCH	43
	INTRODUCTION	43
	RESPONSIBILITIES OF THE APPLIED BEHAVIORAL RESEARCHER	43
	Scientific Requirements	44
	Practical Requirements	44
	Real-World Considerations	45
	PURPOSES OF APPLIED BEHAVIORAL RESEARCH	46
	Hypothesis Testing	46
	Problem Exploration	46
	Evaluation	46
	APPLICATION CONTEXT	47
	Older Systems	47
	Newer Systems	48
	Contrast of Older and Newer Systems	48
	Implications for Applied Behavioral Research	49
	RESEARCH VEHICLES	49
	Real World	50
	Simulation	50
	Laboratory	50
	Comment	51
	FIDELITY AND VALIDITY OF RESEARCH SIMULATORS	51
	Fidelity	53
	Validity	54
	CONCLUSION	

Appendix		Page
D	FIELD INTERVIEW – MARITIME BEHAVIORAL RESEARCH – <i>Dr. Thomas J. Hammell and Dr. Myriam W. Smith</i>	
	SCENARIO CONSTRUCTION	55
	REALISM OF SIMULATION AND SCENARIOS	55
	SUBJECTS	56
	PERFORMANCE MEASUREMENT	56
	REDUCED SIMULATION VS. FMS	56
	USING SMEs	56
	TECHNIQUES FOR CONTROLLING FMSs	57
	COMPARABILITY OF STUDIES	57
E	FIELD INTERVIEW – NUCLEAR POWER PLANT SIMULATION-BASED RESEARCH – <i>David Woods</i>	59
F	FIELD INTERVIEW – THE BOEING COMPANY – <i>Thomas C. Way and Dr. Richard E. Edwards</i>	63
	APPROPRIATE LEVELS OF SIMULATION FOR RESEARCH AND EVALUATION	64
	SCENARIO DEVELOPMENT	64
	CONTROL OF SCENARIO	65
	DATA COLLECTION	65
	ATTITUDES OF PARTICIPANTS	65
	EDUCATION AND RESPONSIBILITIES OF THE SIMULATOR RESEARCHER	66
	ATTITUDES OF CUSTOMERS AND USERS OF INFORMATION FROM SIMULATOR STUDIES	66
	PLANNING FOR DEMONSTRATIONS	67
G	FIELD INTERVIEW – SIMULATION STUDIES OF VISUAL ILLUSIONS DURING NIGHT APPROACHES – <i>Dr. Conrad L. Kraft</i>	69
H	FIELD INTERVIEW – SIMULATION AT THE UNITED AIRLINES FLIGHT TRAINING CENTER – <i>Dave Shroyer</i>	
	LOFT TRAINING	71
	CLR TRAINING	71
	COMMENTS	71
	Problem Selection	72
	Scenario Construction	72
	Scenario Testing	72
	Instructor Training	73
	Cabin-Crew/Flight-Crew Interactions	73
	Stress and Peer Pressure	73
	Organizational Pressure	73
	Nonpilot Observers	73
	Fidelity	74
	Validity of Behavior in Simulators	74
	LOFT Guidelines	74
I	FIELD INTERVIEW – McDONNELL DOUGLAS CORPORATION MILITARY MISSION-ORIENTED SIMULATION RESEARCH – <i>Dr. William J. Cody</i>	
	CUSTOMER RELATIONS	75
	SCENARIO DEVELOPMENT	75
	SIMULATOR ISSUES	75
	SUBJECT ISSUES	76
	Selection and Training	76
	Physical and Psychological Well-Being	76
	Subject Attitudes	77

	Page
Appendix	
I CONTINUED	
PERFORMANCE MEASUREMENT	77
SCHEDULING	78
BIBLIOGRAPHY	79

CHAPTER 1

INTRODUCTION

PURPOSE

The purposes of this study were to:

1. Examine the state of the art of mission-oriented simulation (MOS) and its use in human factors research,
2. Discuss the issues involved in determining the level of fidelity that is appropriate for the conduct of behaviorally oriented human factors research in civil multicrew transport operations,

3. Develop guidelines for the conduct of behaviorally oriented human factors research in civil multicrew transport operations that require full-mission simulation (FMS), and

4. Develop recommendations for future research which might fill gaps in present knowledge regarding the validity of simulation research and suggest alternative methods which might improve the productivity of such research.

Although many of the principles we discuss have wide application, this study was concerned only with the simulation of civil air transport operations. In this context, FMS includes simulation of all of the stimuli presented to the flight crew. It includes the aircraft cockpit, visual and motion cues, aircraft flight dynamics, all of the aircraft subsystems, the flight environment (including air traffic control (ATC), weather, and other air or ground vehicular traffic), the cabin crew, and all ancillary flight services (such as dispatch, ramp passenger services, and maintenance). Full-mission simulation lies at the high end of the range of fidelity associated with MOS. The more restrictive term "line-oriented simulation" (LOS) refers to the MOS of a civil airline operation.

PROBLEM STATEMENT

The problem addressed was the identification of effective methods for developing and using MOS and scenarios that represent the operationally complex environment required for human factors research in civil air transport operations.

A primary goal for the human factors researcher is to produce experimental conditions that elicit behavior that would occur under similar circumstances in the real world. The ultimate consideration is *performance in the real world*. The importance of this concept in the practice of human factors was stressed by David Meister (1985, p. viii) when he wrote:

...the purpose of ergonomics/human factors is to describe, analyze, measure, predict, and control the real world of systems functioning operationally (i.e., not under experimental control).

The obvious implications of this statement for human factors researchers, again in Meister's words (1985, p. viii) are:

...in consequence, the ideal environment in which to gather data is the operational environment. It may be necessary for various reasons to measure in some environment other than the real world, such as a laboratory or a simulator, but in such cases the conclusions derived from the data must be verified in the operational environment.

PROCEDURAL APPROACH

Our first step was to review the approach and the relevant literature, and to conduct field interviews with recognized experts. Next, we analyzed the information we had obtained and discussed it with authorities in the use of mission-oriented behavioral research. Our final step was to write this report.

LITERATURE REVIEW

The existing literature was screened for information related to MOS and the development of scenarios for behavioral research. Although NASA has used MOS for training since the beginning of the space program and major airlines began exploring the training potential of FMS in the late 1970s, little has been reported on its use for human factors research.

Few FMS studies have involved civil aviation operations. One of the few was the pioneering study conducted by the NASA Ames Research Center in the mid-1970s (Ruffell-Smith, 1979). Two more FMS studies were being done by researchers at NASA Ames while this study was being conducted (Foushee, appendix A, and Murphy, appendix B).

There also was a paucity of studies dealing with the predictive validity of results achieved using MOS for research. Most of the reported studies were concerned with the performance in a simulator alone, or with the transfer of training. The transfer-of-training studies were not particularly helpful because, while they can predict the amount of simulator training that will contribute to real-world task performance, they do not necessarily reveal the real-world validity of the behavior that occurred.

Unquestionably, there are common elements among simulation requirements for training and for research. However,

there are important differences. These differences arise not only because research objectives and training objectives are different, but also because behavioral research imposes more stringent research requirements for the control, measurement, and generalization of results. With the exception of the publications of Lauber and Foushee (1981) dealing with the use of FMS in flight crew training, we found that the important issues of scenario development for FMS have been given only cursory attention in the literature.

Despite these difficulties, the literature search was helpful. While training was the purpose of most of the reported MOS and LOS work, the training literature addressed many scenario issues that are relevant for the design, development, and execution of simulation scenarios for research.

For example, the training literature stressed the importance of a multidimensional concept of fidelity. "Fidelity" has been described in many ways. It defies simple description or measurement and, unfortunately, there are no easy answers to questions of fidelity requirements for research (see appendix C). It also is clear that there are controversial issues regarding the role of physical fidelity in simulation — considering such areas as the importance of platform motion, low-amplitude vibration, realistic control loading, and visual system fidelity. Here again most studies have dealt with the effects of these variables on performance in the simulator or on transfer of training. The predictive validity of a simulator performance on real-world flight operations is far from clear.

FIELD INTERVIEWS

Field interviews with experts in the use of MOS were particularly helpful. Time and budget constraints prevented our contacting all knowledgeable people in each area of interest; and in some cases we were restricted to telephone interviews. However, because we were able to selectively interview recognized leaders who were well informed regarding activities in their respective fields, we were able to take advantage of the insight and practical knowledge of many of their fellow researchers.

REPORT SCOPE AND ORGANIZATION

This report is concerned solely with the use of MOS for human factors research in civil air transport. Its intended audience ranges from qualified researchers who may not be fully familiar with the FMS of the air transport environment to corporate or institutional research managers who have to make difficult judgments regarding which research programs to support.

The areas discussed range from the myriad of conceptual issues which have to be considered and evaluated when plan-

ning to the pragmatic lessons learned from past FMS studies. At the onset of the study, we were asked to concentrate on the appropriate use of FMS, on considerations of scenario design, and on their implementation for research purposes. We were asked to avoid such other issues as requirements for motion and visual systems.

While this report assumes basic training in behavioral research, including familiarity with experimental design and statistical methods, more than traditional training is required for MOS research. A critical domain of information lies between the lessons learned in basic behavioral research training and the requirements for effective FMS research in aviation. It is information that has been learned on the job by those researchers who have had to perform applied research in complex simulation environments.

As expected, clear answers to the critical issues were not always available because often there was a conflicting interdependence among considerations that outwardly seemed to be well-defined and separate. Resolution of pragmatic problems which were created by the desire to study behavior in a simulated environment usually required reasoned trade-offs and judgments by highly skilled researchers. It was not always possible to reduce such judgments to a firm set of principles and procedures. Moreover, in many cases, time and budgetary limitations dictated the boundaries of what could be considered in a simulation study. Despite these difficulties, our goal was to identify the basic issues involved in planning applied behavioral research utilizing FMS, discuss the wide gamut of considerations that must be evaluated, and provide guidelines for the use of FMS.

Chapter 2 discusses the fundamental issues and requirements of applied behavioral research for which MOS is the research vehicle. These issues include problem definition, study-plan development, experimental strategy, test preparation, data collection, data analysis, and interpretation of results. Each is important, and each requires specific consideration in all behavioral simulation studies, whether they involve part-task or FMS.

Chapter 3 then narrows the scope of discussion to the development of guidelines for doing full-mission LOS research. Topics include the foundations of LOS, when it should be used for research, research team composition, research subject selection and training, scenario development, scenario script writing, operating team training, scenario testing and running, subject debriefing, and lead-time considerations. A preimplementation checklist as a summary of the chapter is included.

Chapter 4 contains research recommendations to fill gaps in knowledge regarding the validity of simulation research and considers alternative methods which might improve the productivity of such research. The recommendations discuss testing full-mission-research validity, studying alternative forms of simulation, optimizing mission-oriented research, developing human-performance models, identifying subjective measures of fidelity, and integrating research efforts.

The appendices are particularly important. Appendix C discusses the fundamental issues and requirements of applied behavioral research. It discusses the general issues that influence how and why applied behavioral research is conducted, the basic responsibilities of the researcher, the three general purposes of applied behavioral research, the application context of the research as it changes with advances in technology, the choice of research vehicles, and the fidelity and validity of research simulations. All of these factors are involved in the rational determination of the research vehicle and environment required for effective, behaviorally oriented, human factors research and is true regardless of whether the requirement is for relatively simple laboratory apparatus, part-task simulation, or a sophisticated FMS.

The field interviews, documented as trip reports, were converted into the eight "case studies" in appendices A, B, and D through I. These case studies range from Conrad Kraft's narrowly focused and classic study of visual illusions during night approaches, through studies involving marine ship handling, and a variety of civil and military aviation applications. These studies offer the insights of experienced

and successful investigators who have had to resolve the issues discussed in the remainder of this report.

The authors are grateful for the time and insights provided by the investigators we interviewed during this study. Noteworthy contributions were made by Dr. H. Clayton Foushee of NASA Ames Research Center; Dr. Thomas J. Hammell and Dr. Myriam W. Smith of the Eclectech Associates Division of Ship Analytics, Inc.; Dr. David D. Woods of the Westinghouse R&D Center; Mr. Thomas C. Way of the Boeing Aerospace Company; Dr. Richard E. Edwards of Boeing Computer Services; Dr. Conrad L. Kraft of the Boeing Company (Retired); Mr. David Shroyer of United Airlines Training Center; and Dr. William J. Cody of McDonnell-Douglas Corporation, St. Louis.

We also appreciated the guidance provided in meetings with Dr. David C. Nagel, Dr. Charles E. Billings, Dr. John K. Lauber,¹ and Dr. H. Clayton Foushee of NASA Ames Research Center. And we thank Dr. Alfred T. Lee and Mr. Robert T. Shiner of NASA Ames Research Center and Dr. Richard M. Frankel of Wayne State University for their thoughtful review of the draft manuscript.

¹Now a full-time member National Transportation Safety Board.

CHAPTER 2

PLANNING AND CONDUCT OF SIMULATION-BASED RESEARCH

INTRODUCTION

This chapter discusses the basic issues involved in applied human factors research, including the application of fundamental methods and the requirements for doing applied behavioral research in a MOS environment. It covers the range of researcher tasks from study planning to data interpretation, and in it the assumption is made that applied MOS research is to be conducted to solve an operational or system-design problem.

Williges and Mills (1982) compiled a catalog of methodological considerations for human factors research in systems. As part of their work, they derived a seven-stage classification for drawing together the methods and procedures applicable to conducting simulation-based research to answer specific system-design questions. The seven stages of planning and execution for human factors research are: 1) Problem Definition, 2) Study Plan, 3) Experimental Strategy, 4) Test Preparation, 5) Data Collection, 6) Data Analysis, and 7) Interpretation.

PROBLEM DEFINITION

The first step in any research program is to fully understand the practical problems to be investigated and the information that will be needed for the user of the research results to make an intelligent decision regarding them. The problem statement must be defined in terms of researchable and testable questions that are operationally meaningful. At this stage, a subject matter expert (SME), although not necessarily working for the researcher, is a critically important resource for developing – or negotiating – the definition of the problem statement. The SME can also be of considerable help in establishing the form of the interpreted results that will be most meaningful to the ultimate user.

Establishing the Operational Need

Frequently the operational user of the research results will frame the research question in practical terms. This can create a very real problem for the researcher because, if the basic problem is not carefully examined, the research question may be framed in terms that provide incomplete or erroneous answers. The research question may be framed in terms that do not include all of the important issues, or in

terms that presuppose the approach necessary to answer the practical problem. For example, the user may ask whether pilot fatigue in a given situation will affect ability to perform normal flight tasks. The issue behind the question can be crew scheduling and duty hours involving not only basic safety issues, but also questions such as the number of crews necessary to sustain flight operations. It is difficult to frame the right question without fully understanding the operational problem; thus, the researcher must learn the nature of the operational problem that led to the practical question to be sure that the research question reflects real-world needs.

Establishing the Research Objective

Once all aspects of the problem are understood, the practical question must be translated into an answerable form. The first step is to determine what relevant information is available. At this point the researcher's job is to secure information – not to do empirical research. Research may be required to answer the operational question, but this should not be an automatic assumption. Frequently the required information can be obtained by a literature review and discussions with others who are knowledgeable in the problem area.

If it is established that additional research is necessary to answer all or part of the practical question, the question must then be translated into testable form. The practical question must be stated as a research question and framed in terms of specific relationships among measurable variables (Cody, 1984). If the factors of interest are not directly measurable, they must be defined in terms of other variables that are. This definition process may again require the use of SMEs to be sure that answering the research question will answer the practical operational question.

Often the practical problem is stated in terms that are undefinable from a research standpoint. The earlier example of a question about the relationship between fatigue and normal flight tasks illustrates the difficulty. Despite a great many attempts, there is neither a widely accepted definition of fatigue (Bartley, 1976), nor an industry consensus on the definition of such basic concepts as cockpit workload.

If the users are not likely to see the relationship between the planned research and the operational problem, a user education program will be needed. This need is important if the operational problem, which has been raised by the user, is not directly answerable. Good communication with the user at this stage of development is the only way to minimize problems of this nature.

In summary, the researcher's main task in the problem definition stage is to collect and organize the relevant facts in order to explicitly determine what information is needed to state the practical question in answerable form. A study plan draft should be prepared that clearly states "what is to be accomplished." At the same time the researcher should be establishing at least a general understanding of the form of the end product, and thinking of ways to state it in terms that are meaningful to the user population.

STUDY PLAN

Developing a study plan includes the following four tasks: 1) frame the general approach, 2) describe the operational conditions, 3) define the variables and develop the preliminary scenario, and 4) perform a sensitivity analysis.

Framing the General Approach

At this stage, particularly if the research involves exploratory or hypothesis testing, the research project should be discussed with others who have different interests and expertise. There can be wide differences of opinion on the optimum approach. Other investigators may see the research problem and approach differently. On the other hand, in the case of evaluative research, fundamental differences of opinion about the approach are less likely. For a discussion of the differences between exploratory, hypothesis-testing, and evaluative research, see appendix C.

The researcher also should be cautious not to let knowledge of the availability of particular research vehicles, especially large simulators, be an overriding influence on what is required to support the research. There always is the temptation for the equipment available to determine what the research vehicle will be.

It is worth remembering that the nature of the research problem may offer an opportunity to structure the research to gain information which has applications beyond the immediate practical problem and contributes to general scientific knowledge. Moreover, a research project that solves a practical problem by developing a general principle almost invariably provides a better answer than research which does not. It has another advantage. If either the characteristics of the system being investigated or the operational problem changes slightly, the results of a study of general principles should remain valid, whereas the results of a study framed in too restrictive a manner may no longer be applicable.

A good example of research that both solved a practical problem and made a substantial general contribution is the series of experiments performed by Kraft (see appendix G) to determine whether characteristics of the B-727 aircraft were contributing to a series of landing-short accidents

during night-landing approaches under visual conditions. His early analysis found commonalities among the external visual scene associated with the incidents. Kraft went on to prove that the underlying reason for the low approaches was not a characteristic of the airplane. The basic cause was a visual illusion of height produced by the combination of lighted and tilted terrain behind the airport and the lack of any lights in the foreground -- a condition frequently found during night approaches over large bodies of water.

The practical solution for the problems associated with these night approaches was an educational program for pilots and a revision of cockpit procedures. If a less skillful researcher had undertaken the research, however, and focused solely on the practical question (i.e., did the characteristics of the aircraft contribute to the accidents?), the answer might well have been equivocal. And even if the practical problem had been solved, the more important general contribution would have been lost.

Kraft's discovery of the underlying cause of this series of accidents led to a better understanding of a fundamental problem. It was a major contribution to aviation safety, and an outstanding example of applied behavioral research. Wickens' (1984) work on a multiple-resources model of human performance and aircraft-design display is another excellent example of theoretically guided research performed at the same time a specific problem was being addressed.

The researcher's conception of the research problem -- as either narrow and specific, or as an example of a broad and general problem -- determines whether the study may have a potential value beyond the immediate solution of the problem. This does not suggest that every applied research project can be expected to contribute to fundamental scientific knowledge. We do suggest, however, that whenever possible it behooves the researcher to choose a research setting and conditions that permit the examination of underlying causes, rather than to focus solely on the solution to an immediate practical problem.

Describing the Operational Conditions

Once a decision about the general research approach has been made, the study plan should be developed. The first step is to develop a detailed description of the operational conditions of interest. Describing the plan can be a simple or complicated task, depending on the researcher's familiarity with the operational environment, the scope of the problem, and its complexity. Describing the operational conditions of interest with precision is an important step and it should be done with considerable care.

The researcher should make every effort to gain first-hand experience in the operational setting. In many cases, an SME may be needed to help develop a description of the operational conditions, determine the equipment to be used, and select appropriate operational tasks. The value of SMEs

should not be discounted, regardless of the researcher's general familiarity with the problem domain. It is obviously important to be sure that the SME is indeed an expert in the critical areas.

An operational description of the study is necessary to identify the factors that must be present to produce the desired behavior. It should include all factors that may modify or otherwise influence that behavior. The operational description should include the measurable manifestations of the desired behavior as well as the operator skills involved. Operational descriptions provide the basis for selecting the independent and dependent variables to be examined, as well as determining subject selection criteria and training requirements if training will be needed.

Quantitative information on the operational factors should be included in the operational description to the degree this is possible. It is important to know the range of values that an operational factor may have. While it is relatively easy to set values for system variables such as aircraft operating characteristics, this is not true for task-performance variables. It is very difficult to develop meaningful, measurable criteria for such task-performance variables as the sequence, timing, and duration of communications or display viewing — or for the interpersonal skills and attitudinal concepts involved in measuring performance in areas such as leadership, judgment, cockpit-crew coordination, and resource management.

Defining the Variables and Developing the Preliminary Scenario

Armed with a general approach and description of the operational conditions, the third step in the development of the study plan is to structure the research project in conceptual terms. This requires the development of a first approximation of the number and values of the independent variables, the conditions to be held constant, and the data needed to derive the desired performance measures. Also at this point, the researcher should start to consider the tasks to be performed, the sequence of events, and when data are to be taken. The performance of this step should result in a conceptual synthesis of the research variables and operational factors that will be manifest in the simulation.

The researcher must determine whether the operational factors and anticipated crew behavior can be expected to produce the variables that he or she wishes to manipulate, control, and measure. For example, the researcher may wish to study decision-making as a function of workload. An operationally relevant task that can be manipulated for this purpose should be a better choice than a task that has no operational relevance. The decisions of the researcher at this stage start the formulation of the preliminary research scenario.

The process of developing the preliminary scenario begins by defining the general mission in concrete operational terms

and then by blocking out its major segments on a sequential or time-line basis. The scenario may be a mission-oriented segment (e.g., take-off or landing), or a full-mission, gate-to-gate flight. The researcher then must define the initial system conditions, environmental conditions, and the effects produced by external agents such as ATC, other traffic, and normal operating events. The special conditions or events representing the test variables also must be included.

The preliminary scenario should then be used to identify subject task and simulator operational requirements. It is important to identify subject task requirements at an early stage so that scenario timing requirements and subject training needs can be established.

The simulator requirements should be defined in terms of required capabilities. These capabilities include the operational and communication subsystems that must be simulated and the environmental effects to be produced. The operational subsystems can include any that are present in the simulated airplane. The environmental effects can encompass such atmospheric phenomena as specific visibility conditions if a visual system is to be used, weather conditions that require weather radar simulation, and cockpit motion to produce turbulence effects. Communications subsystems that may be needed include those for ATC and company communications, cabin and ground crew intercommunication, and a passenger address system.

Scenario development is an iterative process. At this stage of the study plan, the preliminary scenario should be developed only to the extent necessary to define how the independent variables will be controlled in the context of operationally relevant tasks; and to determine subject, task, simulation, and data collection requirements. Further development of the scenario should continue during test preparation.

Performing the Sensitivity Analysis

After developing the preliminary scenario, the researcher should do a sensitivity analysis (Cody, 1984) by thinking through how the conditions, events, and tasks are likely to influence the behavior of interest. A sensitivity analysis requires reconsideration of each operational factor to determine if it is relevant and necessary to include in the simulation, and to estimate the magnitude of the factor's effect on the desired behavior. In some cases the effect may be too little to reliably measure, or so great as to be disruptive. In either case, the study plan may have to be changed.

The researcher must examine the preliminary scenario to discover possible deviations from the expected sequence of behavior that would complicate data analysis and interpretation. Consideration of this issue is crucial in line-oriented FMS because of the decision flexibility that line flight crews have in their day-to-day operations.

Much of the information derived from the sensitivity analysis will be useful in constructing the formal design of

the study, including estimates of the number of runs required, the number of subjects needed, and the amount of data to be accumulated. All of these issues have practical, as well as scientific implications, for they directly affect study costs.

EXPERIMENTAL STRATEGY

Experimental strategy includes selection of an experimental design, assignment of subjects to treatment conditions to control variability between subjects, and also includes the development of performance measures. Here we are departing slightly from the order suggested by Williges and Mills (1982). In their seven stages of research, they placed experimental strategy after test preparation. Because experimental strategy, experimental design, and the development of performance measures affect the planning and execution of the research, we believe experimental strategy should be considered at this stage of the process.

Experimental Design

In most human factors research studies there is a conflict between the real-world complexity of the operational tasks and the need for economy of the research effort. Neither time nor money are unlimited. There is also a conflict between the need to elicit behavior that is equivalent to the behavior that would occur in the real world (and therefore the behavior necessary to produce meaningful data), and the need to exercise control to minimize variability, therefore maximizing the opportunity to discover reliable and statistically significant effects. A good experimental design strikes an appropriate balance between these conflicting demands.

Williges and Mills' (1982) *Catalog of Methodological Considerations for Systems Experimentation* discusses the advantages and disadvantages of several experimental-design alternatives for applied behavioral research. Their discussion focuses on the kinds of situations that arise in simulation-based systems research, and the trade-offs among design alternatives in terms of the data collection effort and the amount (and reliability) of the information that can be obtained when rigorous standards are applied.

Cody (1984), *Recommendations for Conducting Manned Flight Simulation Research*, reviews the major issues raised by Williges and Mills and then goes on to discuss the practical implications of choosing certain design alternatives. For example, a sequence of small studies, whether combined in a full or fractional factorial design, requires more calendar time to complete than a larger study that combines them. Therefore it creates more opportunities for nuisance factors such as subject drop-outs or simulator failures to interfere with the research project.

Researchers who have received a traditional education in full-factorial designs will find both of the referenced reports to be useful overviews of the range of design alternatives available, and of their scientific and practical implications for applied behavioral research.

Performance Measurement

Performance measurement of the human and of the system being simulated is needed to answer research questions. In this case, measurement implies more than simple data collection. In cases in which there is a standard or criteria to allow interpretation of a measure in terms of real-world performance implications, it is usually in the form of a summary index (or several indexes), as a result of the aggregation of data from the same or multiple sources. If there are no established criteria for performance, the lack of criteria will greatly affect the conditions of the research, the types of data collected, and the number of measures developed. The prudent course is to maximize differences in conditions and to have enough measures to be reasonably sure that some measures of performance will show significant differences.

The development of performance measures begins with a definition of the behavior of interest in terms of detectable events. Frequently, the performance itself may be the focus of concern. Studies to determine what a person, crew, or system can do are usually interested in observable, overt performance. For example, the question may be, "how long does it take a pilot to respond to a warning indicator and begin corrective action?" In this case the performance measure is defined by the question.

If the behavior of interest is covert — as it is in a decision process — developing a measure is much more difficult because overt manifestations of the decision process may be no more than pressing a button, or taking no overt action at all. This type of measurement problem is becoming common with the shift in research interest from what *operators can do*, to what *supervisors do* in existing systems, and what they *will do* in new systems.

The magnitude of the performance measurement problem is proportional to the complexity of the simulation. This is because all of the factors that influence the behavior of interest must be included in the simulation to elicit realistic behavior. The less the researcher knows about the relevant behavioral processes and the operational factors that affect it, the more likely it is that he or she will use a complex and comprehensive simulation. There is less certainty of what to measure if the researcher is plowing new ground.

Unfortunately, if a large-scale simulation is required for the study, a difficult and complex effort will be required for developing methods of measuring performance. There is some guidance in the literature on crew/system performance measurement (c.f. Vreuls et al., 1977, 1985a, 1985b), but it is not complete. There are substantial criterion issues and,

in many cases, there is not a theoretical basis for measurement.

The best course of action is to examine what has been measured in previous studies, listen carefully to SMEs and, if possible, discuss measurement issues with other investigators who have dealt with this problem for similar tasks. These activities should define candidate measures, which should then be tested empirically to determine if they produce performance differences which are detectable by other means, such as (but not limited to) ratings by expert observers.

TEST PREPARATION

At this stage, the researcher should have a well-developed study plan and can now start thinking seriously about the many practical matters that will need attention prior to the actual conduct of the research. The practical matters include determining simulator capabilities, resolving any conflicts between simulator capabilities and research requirements, coordinating with simulator support personnel, developing the detailed scenario, establishing procedures for the conduct of the experiment, recruiting subjects, training subjects, and pretesting all equipment and procedures.

Simulator Capabilities

In most cases, simulator facilities are constructed to support flight crew training, engineering studies, or broad, long-range programs of research. One of the researcher's first tasks is to discover the specific simulation capabilities that are available, and to determine whether or not modifications can be made if they are necessary for the proposed research.

An early visit to the simulator facility may be sufficient to discover if there are conflicts between the research requirements and the simulator's capabilities. This initial visit should be a prelude to several meetings with facility personnel to gain a thorough understanding of the simulator features, its operation, and the duties of each member of the simulator support staff. It is important to establish a schedule of meetings to work out the details of preparing for the study and a schedule for preparing the simulator. Individual responsibilities of the personnel involved should be determined at this time.

Conflicts Between Simulator Capabilities and Research Requirements

Conflicts between simulator capabilities and research requirements usually involve details rather than gross discrepancies when modern simulators are being used. For example, the researcher may want a particular event to occur contin-

gent upon the occurrence of one or more other events. While the simulator may be capable of producing the event desired, it may not be able to produce the event at the desired time or under the desired conditions. This sort of conflict may not be discovered until the final scenario is established, and the programming and other work is under way.

It is impossible to list all of the possible problems associated with the capabilities of the simulator that may be encountered. Many will be discovered during preliminary testing; others may not be discovered until the actual data collection. To discover as many problems as possible, the researcher should become familiar with the operating details of the simulator early, and should plan upon spending considerable time with the simulator support personnel reviewing details of the research requirements.

A very common problem is whether to modify the research program or the simulator. At the conceptual level, the research program is normally within the control of the researcher. However, the simulator can be under the control of a higher level of management, or another division of the organization. While it is usually easier to modify the research plan, in some cases modification of the simulator may be necessary. Unless the simulator characteristics are known well in advance, and the need for the modification anticipated, the cost and time required to modify the simulator can be substantial. An additional complication may be that the simulator is also being used for other research or training, the needs of which are not compatible with the proposed modification.

Coordination with Simulator Support Personnel

Experience has shown that the researcher cannot always assume that a single person will be the key simulator facility contact to coordinate all aspects of the preparation, scheduling, and operation of the simulator. Usually several people will be involved and the researcher must be prepared to coordinate their activities. The simulator support personnel will not know what is important to the research project or what must be done, unless it is spelled out for them. (See Way and Edwards in appendix F.)

The researcher may have to work with several specialized support people, and should make every effort to understand their responsibilities — the things they can and cannot be expected to do. If physical equipment needs modification, equipment engineers and technicians will be involved.

At this stage, it may be necessary to consult as many as three programmers regarding varying aspects of computer control. Normally, a real-time programmer will be responsible for the software controlling the simulation, the sequence of events, and the control that can be exercised at the console. If a computer-generated-image visual system is used, a visual-scene data-base programmer will be needed to perform any changes or additions to the visual scene. Changes in

visual scenes are described easily in terms of objects to be portrayed, visibility characteristics, and visual events, but even simple sounding modifications can take days or weeks to program.

A third type of programmer may be needed for data acquisition, if there will be event- or time-contingent sampling, or processing of data between capturing the data and recording it. A special programmer for data analyses may also be needed subsequent to the test runs. The researcher and the last two types of data programmers should jointly plan the data collection and analysis methods. Ease of data analysis can be enhanced greatly by the way data are labeled and organized when collected.

The researcher should know enough about the console to operate the simulator if necessary. This means that he or she will have to become acquainted with the operation of the control console, and will be able to set initial conditions and invoke certain options. This is particularly important for checking that the right conditions have been set for different trials. It is very easy to set the wrong sequence of trials if multiple sequences with different conditions are involved in the simulator runs.

Various technicians will be responsible for operation of major subsystems, such as the motion base and the visual system. The researcher should have at least a general familiarity with the duties these people perform, because should a fault occur, he or she would know who to turn to for help.

Scenario Development

The scenario creates a simulated real world for the flight crew, and creates the events of the flight to be observed and measured for the research purposes. The scenario establishes the mission objectives, tasks, environmental conditions, event timing, and rules of operation.

As we have discussed earlier, scenario preparation should begin while the study plan is being developed. Because simulator capabilities are critical, some modification of the scenario may be necessary. At this stage the researcher should be concentrating on refining the scenario to be sure that it is consistent with actual operations. Specific identities, values, and occurrences need to be assigned to the previous, more generally defined characteristics of the scenario. One or more SMEs (e.g., an experienced crew member for the type of aircraft simulated, or an experienced controller for the ATC areas) will be needed to establish these details.

If the same subjects are expected to participate in more than one test run, the researcher may need to develop variations on the scenario to create apparently differing test conditions. The problem is to maintain scenario equivalence at some level of conception for all conditions considered to be constant, even though those conditions may appear to the subjects as different. Cody (1984) was able to overcome this problem in a study of low-level navigation and attack missions of military aircraft.

Cody's flightpath involved several waypoints and multiple targets. He constructed a template of the full flight route and rotated it several degrees for each variation of the scenario. In effect, all of the relative values of flight-segment length and turns remained the same, but the absolute values of all headings changed. The flight legs were long enough that the pilots did not see the constant pattern in the various missions, therefore each scenario appeared to be unique.

Woods (1984) suggests that if the class of behavior of interest is well defined, many specific instances of members of that class can then be regarded as equivalent. For example, it is possible to create a variety of decision problems as long as they meet the criteria defined for a class of decision problems. If this rationale is to be applied successfully, the researcher must have a valid theoretical basis for defining the behavioral class, and the simulated examples of the behavioral class must have operational relevance.

This line of reasoning again argues for the importance of translating practical problems into a more theoretical context to seek generally useful solutions. However, if the nature of the practical problem excludes any basis for establishing equivalence between expected behaviors, the researcher has no choice but to use new subjects for each variation in the scenario.

Procedures

The execution of any MOS is complex. Extensive research and support team coordination, teamwork, and practice are required to execute the simulation properly. In this context, "procedures" are the organizational, management, and operational plans and schedules for the data collection.

The procedures should outline the responsibilities for each of the members of the research team and support crew, and should include contingency plans for potential problems, such as a simulator failure. Logistic needs, such as preflight briefing materials and training rooms and equipment, should also be addressed in the procedural plans. A checklist for the initial state of all simulator switches and control settings is a key procedural document.

The subject testing schedule, the training subjects are to receive, and the amount of practice to be permitted are an important part of the procedures (Williges and Mills, 1982). The training programs and the performance criteria to be met at the end of training must be developed, and the supporting documentation (e.g., questionnaires and forms for keeping a log of the test runs and for recording observational and subject biographical data) should be prepared as part of the procedures. If collateral testing of the subjects is part of the research program, the protocol to accomplish it should also be developed at this stage.

An overall schedule for the pretesting and data collection phases will need to be prepared. Time allowances in the schedule should also be made to include the daily checking

of the simulator, the potential time loss caused by equipment failures, rest periods, and additional, perhaps unplanned, trials to be performed.

Finally, external requirements should be considered in the master schedule. If a research project is the subject of high-level management attention within, or external to, the organization or, if the research vehicle is inherently interesting (e.g., a full-mission simulator), there will be a requirement for demonstrations and tours. Unless these visits are controlled, they will be scheduled to fit the availability of the visitor rather than the convenience of the research.

Unexpected interruptions can occur and often have a disruptive effect, particularly if experienced operators are used as subjects and may not be available at another time. The best course is to fix specific days or times of day for tours and request support from management in arranging tours or visits at those times. This must be done well in advance of data collection periods. Managers should be kept informed of the research schedule to provide them with an opportunity to diplomatically arrange the schedule of important visitors so as not to inconvenience the visitor or the progress of the research.

Data Collection System

Throughout this chapter, the research project has been discussed as if there was only one behavior of interest. This was done for the convenience of discourse, but is unrealistic. In most cases the researcher will be interested in several behaviors. Invariably, a large array of system performance data will be collected and both physiological and behavioral data may be taken. There will usually be several different methods of data collection.

Some data will be recorded automatically using the simulator, other data may be derived from observers, or through video and audio tape recordings. Still other data may be obtained by questionnaires and interviews. All of this creates a data management problem. A good rule is to automate as much of the data collection as possible. The researcher should be aware of the advantages of automated data as well as of its inherent limitations, particularly for behaviorally related research.

At least two actions can be taken to minimize the data management problem. The first is to carefully evaluate what data are required to answer the research and practical questions, and reject any data collection that is unlikely to be helpful. This applies not only to particular kinds of data, but also to the sampling rate. For example, there is no point in collecting control-movement data at a rate of 100 Hz since it is more than a full log unit above a human's response bandwidth.

The second action the researcher can take is to devise a system consisting of procedures, facilities, and computer programs for processing and organizing the data from the

time of collection to final analysis. Cody (1984) describes the structure for one such system that involves five phases: 1) data collection, 2) initial data reduction, 3) single-run result summaries, 4) data base accumulation, and 5) statistical analyses. The development of such a system is well worth the considerable effort it requires. A researcher can be overwhelmed with data from a simulation and it is nearly impossible to assemble the data in a meaningful way after it has been collected if the data analysis methods are not given special attention at each stage of data reduction and analysis. Typical problems include determining how the data are related to the problem and how the data will be structured.

Subject Recruitment

Subject requirements for applied behavioral research in nonmilitary settings usually fall into two major classes — totally naive or highly experienced. In the military environment, personnel with a range of training and experience are often available. This is not always true in the civilian world. Subject experience or skill requirements depend largely upon the scope of the tasks to be performed, the requirements of the ultimate user, and whether the simulated system exists or is a new concept.

In most cases, experienced subjects will be required for simulation studies of aircraft operations. Naive subjects can be appropriate for laboratory studies if the tasks are simple and training requirements are minimal.

A dilemma occurs if the simulation involves rather dramatic changes from a conventional system. There may be enough similarity to the conventional system that experienced subjects would appear to be a natural choice. However, in this case old habits and preferences may unduly affect the subject's behavior (Sheridan and Hennessy, 1984). On the other hand, naive subjects may require extensive training.

Finally, it is always possible that "seasoned experience," which is missing in naive subjects, is desirable because there is the real-world probability that experienced airmen will be the ultimate users of new technology, and if there is a conflict with existing procedures or habits, or negative transfer effects, it would be advantageous to know this at the research stage of the new technology. There is no good general answer as to which type of subject would be most appropriate. The choice can be made only on a case-by-case basis.

If experienced subjects are necessary, the researcher should start recruiting well in advance of the scheduled testing. It is always difficult to find experienced subjects. In addition, because the researcher cannot expect to have experienced subjects available for extended periods, it is important to minimize the amount of time that is required per individual subject.

Delivering materials to the subjects in advance is one way to minimize both the drop-out rate and the amount of time for which both naive and experienced subjects are required.

The advance materials should include a comprehensive briefing on the project, a description of what will be expected from the participant, and training materials to prepare the individual to the extent possible. Questionnaires or other forms that can be filled out in advance should be included (Cody, 1984). The introductory package can motivate the person to continue to participate, and can reduce the use of valuable on-site time for briefings, training, and administrative matters.

Subject Training

It is necessary to provide subject training or familiarization prior to the test runs in most studies. Obviously, this training will vary in scope, depending upon the background of the subject and the nature of the research tasks. Once the training needs and procedures have been established, a principal remaining task is to establish a criterion for training completion.

Intersubject variability is desirable in behavioral research that requires subjects with a range of skills and knowledge, e.g., when individual or crew differences are the object of the study. However, such variability unduly complicates other kinds of research. In many cases, it will be highly desirable to reduce normal intersubject variability to minimize its confounding effect upon the analysis of the data collected. In these cases, all subjects should be equally prepared to perform the research tasks. The criterion may be either a performance test or the completion of training. If completion of training is defined as a certain amount of practice time, invariably it will be accompanied by varying levels of competency. A practical criterion is having one or more successful completions of the same (or nearly the same) scenarios that will be tested. If the training is a practical exercise, data taken during or at the end of the training can be used as collateral data.

Pretesting

The formal pretest is a dress rehearsal for actual data collection and is an absolutely essential step for either a part-task MOS or a full-mission LOS. The pretest involves a complete check of all equipment and procedures. The testing of individual scenario elements should occur throughout the development period as each one is completed. The individual elements must then be combined in a smooth and operationally logical fashion. The research team should step through the whole process several times before the formal pretesting. This is the time to discover any obvious shortcomings in the procedure, and also to develop skill in performing individual tasks. When the scenario can be run smoothly and reliably, individuals with the experience and skills equivalent to the test subjects should be used for the pretest.

The pretest should cover the entire period from subject arrival to exit from the premises. Pretest will be the first opportunity for the research team to test the procedures they expect to use in data collection. The experiences of the research team, and the solicited comments of the pretest subjects should be used to identify any residual scenario problems because pretests rarely go perfectly. The amount of research team practice required to run a perfect scenario is often underestimated because there is a tendency to believe that everything will be in order the next time through, so no additional pretesting is considered necessary. Experienced researchers do not make this assumption, and will pretest until at least one successful run has been completed.

In addition to its verification and training value, pretesting gives the researcher a first look at the effects of the manipulated variables and other test conditions on the behavior of the subjects. The appropriateness of the levels of the variables can be assessed at a time when it is possible to make changes. Pretesting also reveals the characteristics of the behavioral data and tells the researcher whether any measurable effects will be produced.

The pretest also provides an opportunity to check the integrity of the data collection system and the quality of the data. More than one simulator study has been compromised by not verifying that the data can be collected, reduced, and analyzed before the actual test had begun.

For example, Hennessy et al. (1981) conducted a study to determine whether unconventional displays could be used to support training of basic, visually guided flying skills. The study involved the teaching of naive subjects to perform two tasks. The first task was to maintain heading, altitude, and course during 90 sec of straight and level flight. The second task was to perform a 360° roll without changing heading, altitude, or course. This task was initiated by the subjects, usually within 2 or 3 sec after the simulator was activated. The simulator was stopped a few seconds after the roll was completed. The roll maneuver itself took about 2.5 sec.

A suitable, automated data collection program adopted from a previous study was used, but was not pretested for this particular application. Later, it turned out that no data for the roll maneuver was collected. It was discovered (after the fact) that the data collection program (because of a requirement in the previous study) did not activate until 7 sec after the simulator was started and the roll maneuver was over before that time. The lesson was clear: real-data collection should never begin until at least one set of data has been collected and run through the entire data-handling system.

DATA COLLECTION

Theoretically, the actual data collection should be routine if the planning and pretesting were done optimally, but this

is rarely the case even with the best of planning. Simulator faults, unanticipated subject behavior, the wrong setting of a condition, or a host of other "gremlins" can upset even the best plans. Additional time and subjects should be allowed to take care of test-run problems.

The most serious problem that can arise is either an undiscovered error in the setting of a condition, or a fault in one of the automated data-collection systems. A mandatory precaution is to run the data through a "quick-look" program on the same or the next day after it is collected to be certain of its integrity and reliability. Otherwise, it may be days or weeks before the problem is discovered. Depending upon the variables manipulated, this check of the data can also confirm that the scheduled conditions actually occurred.

It is most helpful if the simulator has the ability to get hard-copy "snapshots" of selected CRT displays. A snapshot of a formatted display, taken at the beginning of each mission trial, showing the status of conditions that are preprogrammed, or set at the simulator console, is a reliable record of the condition settings.

Individual subject differences, i.e., those unaccounted for by study factors, almost always contribute more to the variance of the data than any other source. Data from a test that measures an ability or characteristic thought to be related to the behavior of interest in the simulation test can be used as a covariate that may account for some of the individual differences. For example, a vision test might be a valuable source of collateral data if the behavior of interest is visual search or the reading of a display. As another example, subject age was found to be the major source of variance in an automated training study (Vreuls et al., 1975). If age data had not been taken and used as a covariate, the independent variables would not have produced a statistically significant performance change. Questionnaire data on experience and other biographical factors could also provide collateral information that can often be helpful in sorting out the test data.

DATA ANALYSIS

The data reduction and analysis procedures should be planned ahead of time to minimize the time and effort required during data collection, and to assure their relevance to the research and practical questions. While most researchers will be familiar with basic data-analysis concepts, if he or she does not have a high level of data-analysis competence, an expert should be used during the planning and analysis of this critical function. The researcher also should be prepared to do additional, unplanned analyses. Frequently, an unexpected facet of the data will suggest additional analyses and may lead to a better understanding of the research problem.

If the research is directed toward system development or improvement, the data are usually analyzed in terms of indi-

vidual, crew, and system performance. Inferential statistical tests can be used to establish the reliability of effects found, and confidence limits on the data. If it is appropriate to the study design, the percentage of variance accounted for is a useful indicator of the relative effects of the manipulated or measured variables and their interactions, and of other analytic concepts.

A recent review by Rouse and Rouse (1984) of almost 200 evaluative studies of complex human-machine systems was performed to analyze the degree to which definitive evaluative results were produced as a function of factors such as the research vehicle used, the type of study, the type of measurement, and the domain of the study (vehicle control, process control, maintenance, and so forth). The authors concluded with this general principle: "One is more likely to obtain definitive experimental results if the method chosen allows a reasonable degree of control, and the measures chosen allow fine-grained analysis of performance."

INTERPRETATION OF DATA

Once the analysis of the research data is complete, a remaining and vital step is to communicate the results to the customer or user community in a form that is likely to resolve the problem that was the reason for the research. If the data are to be used for systems design, the results should be in a form that permits engineering trade-off comparisons. Other applications may call for general, or composite presentations (Williges and Mills, 1982).

If the study was able to produce findings that can be generalized to a class of problems, or is a contribution to a body of scientific knowledge, the researcher should publish the results. Regardless of the outcome, if the study involved a large-scale simulation, the researcher should describe the planning and execution of the study, the resolution of any problems encountered, and the lessons learned. The lessons learned should include both the things that went particularly well and the things that didn't — i.e., those things that the researcher will do differently next time.

CONCLUSION

This chapter has discussed the process of applied behavioral research in MOS. The research process has been well organized by Williges and Mills (1982), and the issues discussed are familiar. Researchers or research managers who are new to FMS research should consult the source documents for a more thorough presentation of these issues than could be presented in this overview. A more fundamental foundation for why and how applied behavioral research is conducted is presented in appendix C for the interested reader.

CHAPTER 3

GUIDELINES FOR LINE-ORIENTED SIMULATION RESEARCH

INTRODUCTION

Performing LOS research, which satisfies both the scientific and operational communities, is a complex and difficult task. Fortunately, many of the difficulties can be alleviated by following the basic principles discussed in chapter 2 and in appendix C.² The purpose of this chapter is to combine those principles with the lessons learned from LOS training and research experience, and present them as guidelines for the use of LOS in applied behavioral research.

Line-oriented simulation is a specialized type of MOS. The mission is a civil air-transport operation. Line-oriented simulation uses sophisticated flight simulators and a detailed and faithful simulation of specific operational segments to provide an environment that has many of the attributes of the real world of civil air transport. When LOS is used for human factors research, the objective of this high level of simulator fidelity is to produce performances equivalent to the performances similar individuals would produce after receiving similar stimuli in real-world operating conditions.

A basic assumption of LOS is that the more faithful the simulation of real-world stimuli, the interfaces between individual crew members, the systems they control, and the systems that influence and regulate their behavior (e.g., ATC, Federal Air Regulations, and company regulations and procedures), the more likely it is that the behavior achieved in the experiment will be the behavior that would be produced under similar circumstances in actual line operations.

A high level of fidelity has a further and practical advantage. It increases the "face validity" of the experiment in the eyes of both the participants in the experiment and in the potential users of the research. This is important because the more face validity the simulation has, the more confidence the ultimate users in the aviation community are likely to have in the experimental results.

A basic difficulty for researchers and users alike, however, is that regardless of the level of fidelity achieved, one cannot be certain that the behavior observed during the experiment is anything more than the behavior in the simulator of that particular individual or flight crew at that particular time. Individual performance varies even under similar conditions in the real-world, and there can be even greater differences

among individuals. As we noted in chapter 2, "individual subject differences...almost always contribute more to the variance of the data than any other source."

A further complication is that, despite significant advances in simulation technology and LOS scenario development, real-world operations cannot be reproduced exactly. The researcher, therefore, can never be certain that the scenario produces all of the cues that might influence the predictive validity of the experimental results. Today, there is little hard evidence to either support or refute the belief that contemporary state-of-the-art LOS produces the same flight crew behavior that would result if the simulated situation occurred in real flight. Significant improvements in all phases of LOS, however, have resulted in an important consensus within the aviation community. The consensus is that not only does modern LOS simulate real-world flight effectively but, more importantly, it produces the equivalent of real-world behavior. The balance of this discussion assumes that the industry consensus is correct.

FOUNDATIONS OF LOS RESEARCH

Despite more than 30 years of Government-sponsored research using flight simulators, and the long history of the military and commercial airlines using simulators for flight training and checking, line-oriented FMS is a relatively new concept. It began with Northwest Airlines' pioneering attempts at what is now known as line-oriented flight training (LOFT), and with Ruffell-Smith's simultaneous use of FMS for research in his landmark study (Ruffell-Smith, 1979) at NASA Ames Research Center. Since then, much material has been published regarding the use of LOS for training. Unfortunately, very little documentation is available regarding its use for research.

Commonality of LOS for Training and for Research

Line-oriented simulation used in research and LOS used in training have much in common. Both share the same general rules for scenario construction and execution. Both stress a high degree of realism and meticulous attention to details to simulate all of the important elements and interactions of an airline operation. Much of this material is derived from LOFT experience.

² Appendix C contains an overview of the more fundamental principles of applied behavioral research that have particular relevance to civil aviation research. It includes a general discussion of the purposes of research, the use of research vehicles, and the central issues involving the fidelity and validity of research simulators.

Without question, the foundation for this chapter is the Lauber and Foushee (1981) *Guidelines for Line-Oriented Flight Training, Volume I*. It is the original and basic text for LOFT. The LOS principles they delineated remain valid today. The LOS principles outlined in this chapter differ somewhat because of their orientation to research rather than to training. They also reflect the results of our literature search, LOS studies which have been conducted since the Lauber-Foushee report was written, the case studies reported in the appendixes, personal communications with recognized FMS experts, and our own experience.

Differences Between LOS for Training and LOS for Research

Line-oriented simulation is concerned with the use of FMS for flight training. Here we are concerned with FMS for research. In training, the objective is to *change* behavior. In research, the objective is to *observe* behavior. While good training scenarios and good research scenarios have much in common, research needs are more rigorous. Frequently, LOS research requires higher fidelity. For example:

A central issue in simulation, whether it be used for research or training applications, is that of fidelity. For training applications, the requirements for fidelity are straightforward in concept... a high degree of fidelity is only useful if it provides an effective training environment. There is no *a priori* requirement for a given degree of fidelity, only for that which will produce the most rapid and long lasting training benefit.

For simulators used for research, the requirements are somewhat different....Here the *a priori* requirements for fidelity are more stringent because, by nature, research is used to explore the unknown.... (Nagel and Shiner, 1983)

Research scenarios also require greater control of crew performance. Deviations from the expected outcome of a training scenario can provide a valuable learning experience. Similar deviations from a research scenario, particularly when accompanied by deviations from the specific outcome or behavior being studied, provide unwanted variance. Unwanted variance can both confound analysis and, in some cases, require a larger number of trials.

A final difference is that when pilots are being trained, they have an understandably closer identification with the outcome of a simulator training session than they can be expected to have with what can be for them a "one-shot" research project. The result can be a tendency for pilots to be more tolerant of fidelity slips in training than in research. This is particularly true if any aspect of the research exercise

becomes uncomfortable, or if their performance becomes unsatisfactory.

When pilots are being trained, they have a high personal interest in the training outcome and can be expected to do everything they can to achieve a successful result. There is nothing approaching this personal identification for a volunteer pilot in a research simulation.

WHEN TO USE LOS

The first steps in any applied research are problem definition, study plan development, selection of experimental strategy, and test preparation. The choice of research vehicle and the decision of whether or not to use LOS methods are inherent parts of this early process. It is worth repeating that simulation fidelity at a higher level than is required can both complicate the study unnecessarily, and make it more difficult to critically examine a narrow research issue.

Some part-task research requires modest real-world fidelity, virtually no scenario, and limited, but highly specialized expertise. Other kinds of part-task research might require a modest scenario of sorts to test a fundamental decision process, or a subsystem's operability or controllability, but little in the way of a real cockpit, or a visual or motion system. Kraft, whose study of visual illusions associated with night landings is one of the classics in the literature, made an important point when he said, "Had a full-mission simulator been made available for the study, ...they probably would have not discovered the basic problem." (appendix G)

At the other end of the scale, research involving interaction of multiple subsystems that are parts of the aviation system can require detailed and carefully scripted scenarios, skilled participants, a multidisciplinary investigative team, and the full resources of today's simulators (Ruffell-Smith, 1979).

All civil aviation systems include a man-machine system which operates aircraft, and an ATC system which controls them. While each is a complete and complex system in its own right, they must be viewed as a coordinated operation in the total system. This report is concerned with the man-machine system, which operates the aircraft, and its interface with a *realistic* simulation of the ATC system. For our purposes, the simulation must be realistic only from the cockpit viewpoint. We are not concerned with simulation of the internal operation of the ATC system.

ADVANTAGES OF LINE-ORIENTED FMS RESEARCH

Line-oriented simulation provides a relatively economical method to observe the performance of people and equipment within the aviation system. It permits observation of the

system's elements (hardware, software, liveware, and environment) under conditions that allow reasonable control of the variables in each element. Without control of those variables, it is difficult to identify the factors responsible for observed performance.

A principal strength of line-oriented FMS research is its unique ability to study the subtle interactions involved in such areas as crew coordination, vigilance, judgment, and resource management.³ LOS can be required if the research involves conditions such as performance degradation, long-duration or infrequent events, or response to emergencies or irregularities as a function of time on-duty. In studies of performance degradation, as might occur in some fatigue states, the desired subject-state often can be achieved conveniently outside the simulator before the simulated mission begins. The conditioned performance must then be evaluated in a full LOS context.

Another advantage of LOS research is that it permits evaluation of the performance of people and equipment as it occurs during transitions from one flight phase or operational mode to another. In some cases, the simulation of a complete flight may not be necessary to accomplish the research objective. It is necessary, however, to faithfully simulate all of the flight phases or operational modes being examined, and to simulate adjacent phases or modes to ensure that any interaction effects that occur can be examined.

LOS also can be used to study a related series of problems involving hardware, software, and behavior in which there are clean breaks between the individual elements being evaluated. An example is a series of related part-task studies involving the evaluation of specific instruments, displays, operational procedures, or controls (appendix F).

When newly developed hardware or software are introduced into the system, its real-world performance can be quite different from the performance that was predicted by its creators, or was observed during part-task studies during development. Performance in real flight, however, is the ultimate criteria — for that is where the performance of hardware, software, and liveware have critical importance. Today, LOS provides the best available vehicle for predicting real-world performance.

In summary, LOS is useful for studying individual or group behavior when the study addresses the following kinds of research:

1. Validation of the results of smaller studies or behavioral hypotheses by observing behavior in a total system context.

2. Evaluation of new or modified hardware, software, or procedures in a systems context before they are introduced into the aviation system.

3. Problem exploration, such as the identification of system problems that occur when individual subsystems are combined into the total system, or observation of the performance and interactions of individual and other elements within the aviation system under the wide range of operating conditions to which they are exposed.

COMPOSITION OF THE RESEARCH TEAM

Once the operational problem and research question have been determined and a decision has been made to use LOS, the researcher's next task is to select a research team. The team's principal tasks will be to develop the scenario and to run the experiment. The following sections discuss the variety of skills and knowledge that will be required.

DEVELOPING THE SCENARIO

The research problem always drives the scenario requirements. This is because the scenario must produce behavior of interest to the researcher and the ultimate user. Because the research must be done in a simulated aircraft and airline environment, the first needs are for expertise in the aircraft and type of operation that will be simulated. Long-haul, short-haul, charter, air taxi, and commuter are examples of type of operation. Equally important is expertise in the geographical area involved and in local meteorological and environmental phenomena such as ice, snow and the slippery runways of winter operations, thunderstorms, valley fogs, wind shears, and the special problems associated with high-altitude flying. Air traffic control plays a major role in air carrier operations and a high level of familiarity with ATC operations, including its vernacular and jargon, is another important requirement for at least one member of the research team.

Meteorological expertise is stressed because weather is an inevitable part of LOS scenarios. The simulated weather must be consistent with the weather encountered in the season and geographical area selected. Most areas of the country have local weather characteristics which have considerable operational significance. Understandably, pilots can be acutely aware of, and sensitive to, them. Considerably more than a general familiarity with aviation weather phenomena is needed to be sure that scenario fidelity is not degraded because of obviously implausible simulated weather.

Even the best simulators have limitations, and specific knowledge of those limitations is required. If the researcher is not intimately familiar with the simulator to be used, and its current status, a person with that knowledge will be needed.

³"Cockpit resource management refers to the utilization of all available resources — information, equipment, and people — to achieve safe and efficient flight operations (Lauber, 1981)." Murphy et al. (1984) define it as: "The application of specialized skills to achieve a crew *organization* and *process* that effectually and efficiently utilizes *available resources* in attaining system objectives."

Finally, a scrupulously detailed script has to be written. There are substantial advantages in having a team member who has had previous experience in writing full-mission scenario scripts. If this is not possible, having at least reasonable access to a person who has written successful simulation scenario scripts can be most helpful.

RUNNING THE EXPERIMENT

Much of the knowledge required to develop a good LOS scenario is also required to run one. Again, it is of prime importance to have people who are familiar with the operation being simulated and its ancillary services. Ancillary services can include cargo, maintenance, weight and balance, ramp service, passenger service, fueling, dispatch, flight operations, and any others that might be involved in the simulated operation. Pilot behavior will be an important part of virtually all studies, and a pilot familiar with the aircraft and its operation will be required. He or she will be needed to deal realistically with the host of minor and largely unforeseen operational problems that will occur, and to provide the researcher with a pilot's perspective regarding them.

A professional air traffic controller, or someone with oral and operational skills very close to one, is a requirement for any scenarios that involve more than an absolute minimum of ATC interactions. This is particularly important because of the virtual impossibility of always closely following scripted ATC communications when one side of the communications loop is entirely unscripted. Of course, there is no way to script the pilot side of the communications, and realistically simulate a real-world operation.

The objection to the presence of observers in the simulator cab is that few real-world airline operations have cockpit observers. There is simply no way that the presence of even silent and unobtrusive observers can enhance realism. However, a very practical reason for the presence of at least one non-crew member in the cab is that in all but the most advanced research models, the simulator's operating console is located there.

There are other advantages to having cab observers. They can record important behavioral and performance data that is otherwise difficult to obtain, and can monitor both the performance of the simulator and the general progress of the scenario. Cab observers can be particularly helpful when the reasons for unexpected pilot reactions or deviations from the scenario are not entirely clear to team members observing from a remote location. Such conditions are bound to occur. The cab observers are therefore in an optimum position to clarify the situation by communicating directly with the team members outside.

Cab observers (and simulator operators) have also served effectively as scenario directors in both training and research simulations when the simulated operational situation

required a flight crew member to leave the cockpit. In these cases, the observer can control the time when the crew member "returns to the cockpit" and can give the returning crew member an appropriate operational message regarding the main cabin or external conditions required by the scenario. This technique assures that the pilot receives an operational message from the returning crew member that facilitates the scripted scenario.

Other important members of the operating team are a simulator operator and a simulator-maintenance technician. The maintenance technician is required because occasional simulator malfunctions are inevitable. Many malfunctions can be repaired rapidly, but this takes specialized expertise. If this expertise is not available, frustrating and expensive delays can occur. Such delays can mean the loss of the entire simulator exercise and, for purposes of the experiment, the loss of a trained and carefully selected flight crew.

Complete and accurate data collection is essential. Depending on the nature of the study, designated individuals may be required for specific data collecting tasks. (Data collection was discussed in some detail in chapter 2.) If such people are required, it is important not to burden them with other tasks which can interfere with their primary responsibility. Even automatic data collecting devices need monitoring. Unfortunately, critical data lost is lost forever.

Finally, there has to be a "wagon boss" — an individual who is usually, but not always, the principal investigator — whose job at this point is simply to coordinate and run the experiment. Prior LOS experience is an obvious help. A LOS is a complex undertaking because regardless of the planning and the preliminary testing, the flight crew may take actions which are unanticipated, the simulator may fail partially or completely, or the research team may make a mistake in controlling the scenario. Any of these problems will require real-time decisions that will affect the outcome of the study and the validity of the results. Planning, training, and leadership are all required to develop a well-coordinated operating team.

In summary, the research team will need positive leadership from the principal investigator and individuals with the following kinds of skills or experience:

1. For developing the scenario:
 - a. An expert in the type of operation simulated and its ancillary services
 - b. An aircraft specialist
 - c. An ATC expert
 - d. A weather expert
 - e. A scenario script writer
2. For running the experiment:
 - a. A person familiar with the operation being simulated and its ancillary services
 - b. A pilot familiar with the aircraft and its operation
 - c. Data collector(s)
 - d. An observer (researcher) in the simulator cockpit
 - e. A simulator operator and technician

- f. An air traffic controller(s)
- g. A "wagon boss"

The preceding discussion does not imply that a separate individual is required to perform each of the required functions. For example, at least one major airline runs a successful LOFT program with a single individual performing all of the roles discussed under "Running the Experiment," except that of the simulator maintenance technician. These individuals, however, are highly experienced pilots and simulator instructors who were carefully selected and trained. It would be a great mistake to expect equivalent performance from people who have not had similar experience.

RESEARCH SUBJECTS

Subject Selection

Major U.S. air carrier pilots are ideal subjects for LOS research. They are familiar with the civil aviation system, and are a significant population of interest. Their training gives them a high level of simulator familiarization and, particularly with the very rapid increase in the use of LOFT, most of them readily accept the simulator as a meaningful replica of their operating world. The same can be true of corporate and regional air carrier pilots, many of whom have considerable simulator experience even though it may not include LOFT.

Little hard evidence is available regarding the relationship between simulator experience and acceptance of the simulated world as the real world. Therefore, if a researcher plans to use pilots with little or no previous simulator experience, and if the research goal is to produce the equivalent of real-world behavior, it would be wise to schedule additional pre-experiment training to ensure a high level of simulator acceptance.

The researcher must always remember that pilot performance will be influenced markedly by the operational practices of the airline that employs them. Although there are nearly universal principles of good operating practice, there are also substantial differences in airline policies and procedures. These differences include critical items such as required call outs, the assignment of duties, and a sometimes undefined expectation of what to expect from other crew members.

If the performance of regular line crews is a research requirement, there are significant advantages in using flight crews from the same airline for the entire study. If that is not possible or desirable, every attempt should be made to schedule pilots from the same airline in the same crew.

There is an obvious caveat. If pilots from the same airline are used for an entire study, the data collected may be representative of pilot performance on that specific airline only, and not representative of airline pilots generally. Although

the jet era has produced more standardization among airlines than was present in the piston era, it would be a gross error to assume that the remaining differences in type of operations, operating philosophy, and procedures among airlines are not substantial.

Performance can be influenced by the type of trips the pilots regularly fly. For example, if short-haul operations were studied with long-haul pilots, the results could be quite different from the same studies conducted with pilots intimately familiar with the intricacies and pace of short-haul operations. The converse is equally true. Wide variations in pilot familiarization with the operation being studied can provide unwanted variability that can be difficult to recognize, manage, or evaluate.

Designers of LOFT scenarios consider these differences, but view them from a training rather than from a research point of view. Their task is to develop scenarios that deal with the operational problems of a specific pilot population, aircraft, procedures, and route structure. As noted by Lauber and Foushee (1981):

The design and development of scenarios for LOFT programs requires considerable attention to the needs of the particular carrier. Different air carriers, different operations within a carrier, and different pilots within an operation all have various types of training needs. It is essential that considerable flexibility be permitted in order to meet these various training requirements.

Similarly, the design and development of scenarios for research will require considerable attention to research objectives and to the impact of subject pilot experience. Differing air carriers, differing operations within a carrier, and differing pilots within an operation can bring a variety of skills, perspectives, and behavior into the experiment. It is essential that these factors be recognized in the selection of the pilots to be used.

If individual performance is of interest, other considerations are total experience, time in type, and proximity to scheduled proficiency training and proficiency checks. The latter is even more important if seldom used flight planning and performance considerations, or abnormal or emergency procedures will be a part of the research scenario. Scheduled training and checking sessions include procedures, calculations, and operational considerations that are not routinely encountered in day-to-day flying and, therefore, need to be periodically reviewed.

Understandably, performance in these areas usually is better just after review than just before the review has started. If the research objective is to study typical performance in routine operations, proximity to training or checking is simply a variable that should be measured and treated as a covariate to understand and account for variations in pilot performance levels that might be related to this factor.

A basic principle in LOS is to minimize the number of things that remind flight crews that they are in the simulator and not in a routine line operation. When pilots are asked to modify or to forget well-established operating procedures, or are asked to fly with crew members who use dissimilar procedures and callouts, it can only serve to periodically remind them that they are not involved in a real-world operation.

The simulator should be configured as closely as possible to the equipment the subject pilots are currently flying. If it is not, a substantial effort for "differences training" may be needed. Such training is always undesirable, but may be unavoidable. In many cases changing critical flight instruments (such as flight directors or course or attitude indicators) to duplicate a specific airline cockpit configuration can produce more realistic and, therefore, better performance data than the best differences training. Changing the simulator configuration might be less expensive than the time and materials required for differences training even if the flight crews are carefully selected and scheduled.

There is another difficulty with training for substantial "differences." Although pilot performance may appear to be adequate under benign training conditions, the basic tendency in all individuals to revert to old and well-established habit patterns under high workload conditions or stress levels is difficult to overcome. Pilots in an unfamiliar cockpit, using unfamiliar procedures, may spend time (and mental capacity) trying to remember how to do something rather than concentrating on what to do. Such reactions during a LOS research exercise might, depending on the study's objectives, produce less than optimum (or even misleading) performance data.

Manuals and forms are equally important. There are wide variations in the design of operating manuals, operational forms, and in other printed material among airlines, manufacturers, and corporate operators. Familiar software of this sort adds realism. Unfamiliar software, especially if it will be used in stressful situations, degrades realism. More importantly, unfamiliar software can also be an unrealistic source of confusion that would not be representative of operations in a familiar environment. It can degrade information-seeking and possibly subsequent decision performance.

Unfortunately, it is virtually impossible to avoid the classic problems associated with sole dependence on volunteer subjects. However, most seasoned researchers who have worked with airline pilots believe this is a relatively minor problem, particularly if the experiment is planned with a generous number of subjects.

Subject Training

Subject training needs, subject training procedures, and the criterion for training completion need to be established if they have not already been determined. Each was discussed in the general discussion of subject training in chapter 2. Here, the additional issues of operational currency and simulator familiarization will be considered.

The importance of operational currency is reflected in FAA recurrent training requirements, and in individual airline policies requiring extra training for pilots who have been away from the cockpit for even relatively short periods. Ruffell-Smith (1979) speculated that even the number of days since the last flight was important to older pilots. While this appears to be a little extreme without additional confirmation, there is little question that pilots who have not flown for long periods do not immediately perform to their usual standard.

If routine operations are being studied, research subjects also need to be current. While retired pilots or naive subjects can be acceptable or even desired in some instances, extra training will usually be required for pilots who are not current. The importance of this step should not be underestimated because pilots cannot be expected to display routine day-to-day performance if that training does not produce the equivalent of full line-oriented familiarization. On the other hand, naive subjects — or experienced pilots who are not current — can be acceptable or even desired in some instances. What is important is that the researcher clearly identifies the research needs and utilizes subjects with the needed characteristics.

Despite impressive technological advances, simulators still do not fly like airplanes. This increasingly minor (but very real) issue is recognized in appendix H of Part 121 of the Federal Air Regulations, which requires recency of experience *in simulators* if they are to be used in pilot training or checking. Part 121, appendix H, also provides specific procedures for reestablishing simulator currency if it has been lost. There is an even greater need for simulator currency if the simulator will be used for line-oriented research.

Currency in both piloting skills and in the equipment being simulated is needed because of the complexity of the flying task and the individual characteristics of modern transport aircraft. One session of differences training cannot turn a DC-9 pilot into a B-737 pilot with equivalent skill. The necessary degree of familiarization with the simulator and the manuals, forms, and trip paperwork required for the study can vary, but should be sufficient to minimize learning behavior once the experiment is under way unless learning behavior is an objective of, or will not unduly confound, the study.

Useful guidance on the amount of training needed can be derived by examining the transition training syllabus of a representative airline for the type of operation and equipment used. The researcher, however, should recognize that these are minimum requirements and that in only a few cases will they produce the equivalent of an experienced pilot operating in familiar conditions.

The timing and coordination of cockpit procedures is a particularly important component of flight crew operations. The required skill appears to decay rapidly with disuse, and minor procedural hesitations can have an adverse effect on an otherwise smooth and professional performance. This can be

an important consideration in LOS research that is dependent on good crew interaction and coordination, especially if the research includes high workload periods.

Recency and currency may not be an issue when new hardware, software, or procedures are being studied. In these cases, a degree of learning behavior is inevitable, as it is during the "shakedown period" that occurs when new hardware or software is introduced into a line operation. Some training and familiarization will be needed. The scope of such training will depend on the research objectives, experimental strategy, and design.

Although periodic testing is a large part of a professional pilot's life, very few pilots enjoy it, and, regardless of protestations to the contrary, pilot subjects are bound to perceive an element of testing in any situation that requires a demonstration of their professional skill. Therefore, it is essential that high-quality training be given, and that high standards of confidentiality and anonymity be maintained. It is equally important that these factors are perceived as such by the pilot participants. Subject egos are important.

THE SCENARIO

"All LOFT scenarios and flight segments should be designed on the basis of a formal and detailed statement of *specific objectives* and desired end products" (Lauber and Foushee, 1981). This principle is even more important in LOS research because of the number of performance options which can arise from a realistic LOS scenario. Performance options are not a comparable problem in training because they can still have a significant training potential. At worst, if the pilot does not follow the expected procedure, it can result simply in a need to repeat the training exercise. If, however, undesired performance options are exercised in research, it will complicate analysis by adding unwanted and confounding performance.

The scenario can be developed as soon as the research objectives have been defined. Scenario development is a surprisingly long and painstaking process which can take considerably more time than it does to actually run the experiment. Subject matter experts in the airline operation and the aircraft being simulated are required. Also, SMEs in areas such as local meteorology, dispatch, passenger handling, and maintenance policies and procedures should be consulted if these areas are, or could become, a part of the scenario.

Details are critical. For example, the weather situation should be consistent with real-world weather patterns that normally occur in that geographical area at the time of the simulation. Pilots will recognize it if the weather is not realistic and judge the scenario accordingly. When elements that will satisfy the basic research objectives have been determined, a time-event-line description of the operational tasks that are required should be one of the first assignments.

A Basic Limitation

It should be recognized at the outset that full simulation of an airline environment is simply not possible. There is no way that an airline flight crew can be expected to drive out to a research institution or training center, climb into a large box-like room which is supported by intricately configured hydraulic cylinders that are surrounded by masses of electronic cables, and not be acutely aware of the fact that they are not about to fly a routine passenger flight.

Fortunately, most airline flight crews are familiar with simulators and have learned to "play the simulator game." They can be expected to become very much involved in the simulator exercise. If they are given a well-designed scenario, they also can be expected to make a good faith effort to react in the same way they would if they were faced with similar stimuli under real-world conditions because this is the way that most of them have been trained and are routinely checked.

Elements of Successful Research Scenarios

Successful research scenarios have included such items as:

1. Sufficient workload to discriminate measurable variables in the performance of interest. (Cockpit workload is a complex and difficult subject which has been defined or considered in a great many ways. It includes decision making; time-sharing and prioritizing concepts; physical and mental tasks; and the control of a wide variety of system-relevant, but not always operationally critical, considerations.)

2. Sufficient time to permit meaningful decision processes and crew interactions. (It is a considerable oversimplification to note that there are at least two kinds of operational decisions. In the first type, there usually is sufficient time to consider the available operational variables. In the second type, critical operational decisions must be made very rapidly, with little, if any, time for evaluation. Scenario designers should be sensitive to the importance of the time-available variable in the decision process.)

3. Assurance that fuel available is a meaningful factor by careful selection of weather, route, and payload. (Varying the time of preplanned holding periods is one effective method of controlling the "remaining fuel" variable. It is worth noting that fuel management has become an increasingly important consideration in contemporary cost-conscious airline operations.)

4. Use of both scenario events and the environment as driving factors in the scenario. (Deteriorating weather, icing, thunderstorms, cross- or tailwinds, wind shear, and wet and slippery runways are examples of environmental elements that have been used successfully for this purpose.)

5. Provision for a number of decision choice points, including the provision of some during flight planning. (Providing decision points during flight planning is a good way to

get the flight crew involved in the simulated action at an early stage. Decisions made during flight planning regarding items such as fuel, minimum equipment list (MEL) items, and other inoperative equipment can affect events and decision making for the entire flight and often will determine the number and kinds of operational variables that are realistically available.)

6. Selection of operational anomalies that will best produce the behavior of interest. (This clearly requires operational expertise in all aspects of the operation being simulated. If the operational experts are to be maximally useful, the researcher must be sure that the experts have a good understanding of the research objectives.)

7. Provision of more than one viable course of action for the crew to facilitate *emergent* behavior, rather than behavior which is controlled rigidly by the constraints of the scenario. (If the study includes observation of flight-crew decision processes and the researcher is interested in emergent behavior, the scenario should include reasonable operational options. Otherwise the only behavior that will emerge will already have been determined by the logical constraints of the scenario. Conversely, if the researcher is interested only in a specific kind of behavior, the number of reasonable operational variables should be restricted. To do this without destroying operational realism requires considerable skill in scenario development.)

8. Inclusion of cabin crew and other ancillary services when they are appropriate. (In many real-flight situations, the cabin crew and ancillary services are, or should be, involved — in some cases only by being kept advised of the progress or events of the flight. Appropriate flight crew/cabin crew dialogue adds a great deal of realism to the simulation.)

9. Selection of performance requirements that are within the ordinary skills of individual pilots and within the skills of an integrated coordinated crew. (The capability of modern simulators to simulate a wide variety of operational irregularities and emergencies can create a temptation for the researcher to complicate the simulated operation beyond reason. This can cause resentment and is an almost certain way to destroy the flight crew's "illusion of reality.")

The Operational Problem

There are virtually no limits to the kinds of operational problems that can be simulated. Problems in hardware, software, liveware, environment, and their interactions can all be studied. Each category can originate from within the cockpit or outside of it. Operational problems can range from relatively simple problems which have no further impact on the flight once they have been diagnosed and corrected, to complex problems which cannot be corrected in flight and have continuing operational ramifications.

Engine starting problems are a good example of a simple problem. After a "hung," or a potential "hot" start has been

diagnosed and properly handled, the engine can be restarted and considered a normal engine for the remainder of the flight. At the other end of the complexity scale, malfunctioning landing- or training-edge flaps, or the loss of one or more generators, electrical buses, or hydraulic systems can affect virtually all aspects of the flight until the airplane is parked at the gate.

Within these extremes, there are many variations and opportunities for a host of imaginative modifications. For example, a loss of pressurization at high altitude can cause an emergency descent which in turn can produce a period of high workload and serious internal and external communications problems. Once the lower altitude is reached, however, the airplane becomes essentially normal except for the pressure loss and a significant increase in fuel consumption. In most cases this will require alteration of the flight plan and a landing short of the destination or planned alternate. It can be a very good way to minimize the number of available airport options.

In one imaginative LOFT scenario, a simulated bomb explosion in a baggage compartment was used to considerably complicate a loss-of-pressurization problem. The simulated explosion resulted in the loss of pressurization. However, the explosion also caused aircraft subsystem problems and unverifiable structural damage. Cabin crew, ATC, dispatch, and emergency services were all involved in the scenario.

Real-World Performance

It is sometimes difficult for the researcher to evaluate pilot behavior in terms of its real-world significance if the level of performance is less than expected, or when there has been an obvious error, including an error in judgment. Understandably, pilots are sensitive about their professional performance. A frequent response, and it can be simply a normal defense mechanism, is to say: "Of course I knew we were in a simulator. If it had been an actual line operation I would have done things differently." It will be virtually impossible for the researcher to know the truth of such statements. Occasionally, some extraordinary behaviors have been involved in air transport accidents.

Evaluating the real-world equivalent performance of subordinate crew members or of the monitoring pilot (the pilot not flying) is another problem. Pilots can also be sensitive to the professional egos of fellow crew members. They, therefore, may be more reluctant to question judgments or to point out errors in a simulated environment than they would in the relative privacy of an airline cockpit. The current emphasis on total crew performance, including recognition of the need for resource management and incapacitation training (which, among other things, stresses the importance of monitoring and full participation of all crew members), helps minimize this problem in line flight operations. A specific

reference to the importance of these operational concepts during preflight briefing (in the context of wanting all aspects of real-world behavior) can help mitigate a reluctance to fully participate during the simulator exercise.

Despite these inherent difficulties, there is a clear consensus among line pilots, instructors, and researchers that something very close to a total line environment can be created. Even more important, they believe that in a realistic simulator exercise pilots become so engrossed in their operating problem that they respond as they would in real flight. There is little question that it is important for the researcher to make all aspects of the simulated exercise as realistic as possible, and to avoid minor intermittent stimuli that jar the pilots back into the world of the simulator with even small cues that are unrealistic.

A Final Comment on the Importance of Scenario Realism

As noted in Cody's McDonnell-Douglas simulation studies (appendix I), the main concern of pilot subjects is mission or scenario fidelity. Pilots do not readily accept deviations from operational practice unless the purpose of the study is clearly to try out new equipment or procedures. Part of their concern with fidelity stems from the fact that a simulation (which is essentially a duplication of an actual mission) is a test of their own capabilities. If pilots are to submit to such testing, they understandably want high levels of fidelity to maximize their opportunity to perform properly. They do not want a shortcoming in the simulation to be interpreted as a lack of personal ability.

THE SCENARIO SCRIPT

Line-oriented flight training experience has demonstrated the importance of detailed scenario scripts. Creating the illusion of the real world requires great attention to detail. To an even greater extent than in training, behavioral researchers need maximum control of performance. It is virtually impossible to achieve an acceptable level of control using a generalized script. The problem is even greater if the researcher is also part of the scenario control team. The additional workload and concentration required by an invariably futile attempt to achieve a realistic scenario from a generalized script leaves the researcher little time to observe the performance being studied.

It is mandatory to script all communications and to use them verbatim. Airline mechanics, dispatchers, cabin crew, passenger agents and other aviation personnel all have their particular communication styles. Virtually any of them can be involved in realistic scenarios. Air traffic control communications are most important. If it is at all possible, a working air traffic controller should be used to provide these

communications. Even then, messages should be scripted meticulously to minimize spontaneous innovations. Spontaneous innovations, while occasionally necessary, are almost always undesirable.

The script should specify the timing of all communications and other elements of the scenario. Each event should be placed on a time/event line which must be scrupulously followed. The script should indicate probable crew responses as well as alternative responses to the extent that they can be predicted. Because the researcher can expect considerable variation in individual and crew performance, it may be desirable to script some kinds of simulated problems by aircraft status or position rather than chronological time. Examples might be when the aircraft reaches a given fuel state (see Murphy, appendix B) or is a specified number of miles or minutes from a geographical fix.

A most difficult problem is to realistically control the number of options that are available to the flight crew without reducing them to the point that the researcher can have no confidence that the scenario is producing the equivalent of real-world emergent behavior. Even when emergent behavior is not required or desired, it is essential that scenario control devices be operationally realistic and tightly scripted. Control mechanisms that have been used successfully in LOFT exercises include the following:

1. Sudden weather deterioration below landing minimums.
2. Passenger service considerations and in-flight passenger emergencies.
3. Runways closed for maintenance, snow plowing, or disabled vehicles on the runway.
4. Bomb threats, or hijack attempts.
5. Subsystem status uncertainties or failures.
6. Traffic delays.
7. Obvious or subtle crew member incapacitations.

These and similar kinds of events can be effective. They also happen in real flight operations.

Scenario control devices should be used with considerable discretion. There is always the possibility that the real message the crew gets when control mechanisms are used is not the scripted message, but the reality-destroying message that in this contrived and make-believe world, the researcher does not want the crew to do something that they would have done in a line operation. If that happens, the inevitable reaction is, "Well we're back in the simulator again." From the crew's viewpoint, losing a viable alternate for an aircraft operational reason, such as the reduced range available because of an engine or pressurization loss, has much more realism than a sudden "truck on the runway." Meteorologically sound weather changes, including changes in winds aloft, have inherent plausibility because of the uncertainty of precise weather forecasts.

SPECIFIC CONSIDERATIONS

Route Selection and Scenario Reality

The illusion of reality is enhanced for the flight crews if the route selected for the experiment is one with which they are familiar (or at least could encounter in their day-to-day flying). Familiar intersections, radio navigation aids, and airways reinforce the validity of the simulation and help maintain the illusion of a line operation. Reality is also enhanced through the use of realistic call signs, including airline names and appropriate flight numbers. Today, airline charter operations are widespread enough to accommodate virtually any research scenario needs, as long as the route segments are chosen carefully and imaginatively.

Most airline pilots are familiar with the routes they fly. Although there are variations among airlines, pilots normally have advance notice of their flights. This permits prior review of approach and departure procedures, special terrain or other geographic considerations, likely routing, general weather patterns, and other relevant factors. If a routine line operation is desired, these are important preliminary considerations in the selection of the routes to be used and the pilots who will fly them. Unless it is contra-indicated by the research objectives, the pilots should know the flight that has been planned for them.

Weather, including turbulence, which is typical of the geographical area and the season, adds a great deal of realism to a line-oriented scenario. Conversely, simulated weather which is not inherently plausible, and this includes all elements of weather, significantly degrades it.

Navigational Aids and Communications Services

All of the navigational aids (NAVAIDS) that are normally on the selected route should be simulated faithfully. This includes providing their proper identifications. If any NAVAIDS will not be available, their absence should be stated in the Notice To Airmen (NOTAMS) which should be available as part of the preflight papers. Any radio-aid identifications which cannot be simulated should be properly NOTAMed. Pilot use of NAVAIDS not required but normally available on the route can be operationally sound (e.g., for double-checking position) and, for the pilot, is an indirect method of checking the validity of the simulation.

Communications to at least three outside sources – company, cabin crew, and ATC will be required. Company communications can involve dispatch, weight and balance, passenger service, maintenance, ramp service, cargo, fuel, gate information, and so forth. These communications vary considerably among airlines, and require careful scripting and familiarity with the operation being simulated. Cabin crew/cockpit crew communications are equally important.

Communications associated with ATC services are complex and will be discussed in the next section.

ATC Communications

Today's airline operations involve an ATC communications contact with a minimum of 11 different controller functions on each flight (e.g., clearance delivery, gate hold, ground, tower, departure, low-altitude enroute, high-altitude enroute, low-altitude arrival, approach, final approach, tower, and finally ground again at the destination). In many cases pilots will communicate with at least that many individual controllers. It can be a nice touch if pilots do not hear the same voice performing each controller function.

This by no means suggests that it would be feasible, or even desirable, to have 11 individuals for ATC communications, but it does enhance realism if the same individual does not simulate all of them. Also, it reduces the possibility of potential, momentary confusion during a hand-off from one controller to the next. The first reaction of a pilot who hears the same voice after changing frequencies to the next controller might be to think that he forgot to change frequencies. This is particularly true in a period of high workload where several actions might be time-shared. Thus, if there are at least two people available to simulate ATC communications, they should alternate, so there is a voice change for each hand-off. To deal with this problem, the NASA Ames Man-Vehicle Systems Research Facility (MVS RF) utilizes an electronic voice disguiser that provides 12 different voices from a single controller to enhance the realism of its simulations.

At some point, background materials such as ATC tapes must be secured, or appropriate scripts developed. They must be typical of the airways and terminals selected, the time of the simulated flight, and the simulated operation. An ATC tape giving visual approaches during simulated CAT II weather, conflicting wind information, or inappropriate clearances to other airplanes can destroy the realism of an otherwise effective scenario. It is important to be sure that all background communications are consistent with the operation being simulated.

There is little "open" ATC communication time during peak operations at busy airports such as ORD, LAX, ATL, LGA, DCA, or SFO. Scripting and then simulating realistic ATC communications at such airports during their peak traffic periods is very difficult. Foushee (see appendix A) has reported considerable success by using taped recordings of actual communications to provide a realistic ATC communications background for busy airports. This is also an effective way to increase scenario realism with the introduction of additional ATC communication voices.

There is a great tendency among pilots to short-cut ATC and other communications protocols in simulator operations.

It is well worth making a special effort to maintain proper communications procedures.

Flight Planning, Dispatch, and Preflight

A continuing task of the scenario manager is to create and maintain an illusion of reality. This requires meticulous attention to the smallest details. As in real flight operations, the scenario should start with flight planning and dispatch. Preflight duties, the cockpit setup, engine start, pushback, and taxi are equally important because these are the items that set the stage. All of them should be carefully scripted.

Interaction with Cabin Crew and Passengers

Simulation of the interaction between the flight crew and its cabin crew and passengers should be a part of virtually all line-oriented scenarios. This is because two-way communications with the cabin crew and public address (PA) announcements to the passengers are an important part of routine operations and an integral part of many abnormal and emergency procedures. If these interactions are not simulated effectively, it breaks the flight crew's "reality chain" any time these communications are appropriate. It is particularly important not to break the reality chain during critical portions of the scenario.

While a basic limitation of modern simulators is that their motion systems require the cockpit to remain sealed throughout the flight, this limitation does not prevent effective simulation of flight crew interactions with the cabin crew and passengers. The effective simulation of these critical elements of good scenarios requires only that the interactions are scripted carefully and imaginatively, and that an operative cockpit-cabin interphone and passenger address system are provided.

The growing number of male flight attendants has made it possible to use either a male or female voice for communications from the passenger cabin. Operationally critical communications, such as those involving emergency evacuations, the whole gamut of cabin emergencies, or problem passenger behavior, can be scripted and add considerably to the realism of the scenario.

If it is appropriate to send a cockpit crew member back to the cabin for a first-hand evaluation of a problem, even this can be simulated effectively by requiring that crew member to get up and leave his or her seat. The apparent return of the crew member to the cockpit should be carefully controlled. The real-world operational effect of the cabin visit can be realistically substituted with a scripted briefing to the returning crew member from the LOS coordinator or other observer in the simulator cab.

When this happens, the illusion of flight is preserved for the crew remaining in the cockpit. The other crew member is

doing what he or she would be doing in real flight (i.e., get out of the seat, leave the cockpit, evaluate a situation, and report back). The overt behavior is consistent with reality and is operationally relevant. Under these conditions, the cockpit workload is usually high, so the obvious physical inconsistencies may pass unnoticed by the remaining cockpit crew. If the scripting has been done well, a positive impression will also have been made on the crew member who left the cockpit.

Pacing, Tempo, and Quiet Periods

The pacing and tempo of scenario elements can play a large role in creating an illusion of actual line operations. While there are occasional high-workload periods, routine airline flights are generally low keyed and relaxed. Emergencies and abnormal situations do occur, but they are rare. It is important that the scenario designer create this general atmosphere if an airline environment is being studied.

The tempo should be consistent with the operation being simulated. Periods of relative inactivity (or quiet periods) should be scheduled as they occur in the real world. Even if this is explained in the preflight briefing, it will be impossible to eliminate the pilot's strong suspicion that the research scenario will involve considerably more than just a routine flight from A to B, and that any quiet period is simply a prelude to an ingeniously contrived flight problem. Usually, of course, this will be true.

Most airline training simulator sessions consist of two 2-hr sessions and are limited to a total of 4 hr. It will be difficult to maintain an illusion of reality for longer periods, particularly those that involve prolonged cruise segments. Long periods at cruise require little pilot activity. These periods can be boring in an airplane, and very boring in a simulator. It is particularly difficult to realistically simulate the cockpit environment of long distance flights. If it is a night flight, the problem is exacerbated.

TRAINING OF THE OPERATING TEAM AND SCENARIO TESTING

The amount of training and indoctrination required for the operating team will depend upon the complexity of the experiment, the skills of the individual team members, and their LOS experience. In addition to being experts in their field, all team members should know the research objectives and the simulator's strengths and limitations. They should have a general understanding of the airline operation being simulated and a detailed knowledge of the scenario and the script.

Special training in flight operations, observer techniques, and the making of value judgments may be required if any of

the specialists are to be used as observers. Dual qualification of the operating team members which permits them to function as backups for other team members is helpful because schedule conflicts and availability issues can be expected during the course of any reasonably long study.

Regardless of the care with which a research scenario has been designed and scripted, it is mandatory to plan a series of test runs or "shakedown flights" to ensure that:

1. The scenario produces the desired test situations,
2. Performance measuring devices produce the required data,
3. Recorders record,
4. Microphones transmit without feedback,
5. All of the myriad details involved are polished and fine-tuned, and
6. The scenario cast demonstrates that they can perform their assigned roles smoothly and realistically.

This is the final step in LOS scenario development and it is critically important; it will take considerably more time than expected. Any but the simplest scenarios will require several iterations, and the researcher should not be satisfied until the entire scenario has been run without any "hitches."

PREFLIGHT BRIEFING FOR PARTICIPATING CREWS

After the scenario has been fine-tuned and the subjects have been selected, a preflight briefing by the researcher is needed to ensure that the pilot subjects understand their role and the purpose of any special training that may be required. This is an ideal time to furnish general operational details such as the flight origin and destination, copies of typical flight plans, weight manifests, and loading forms.

It is important to give subjects only general information that will not reveal parts of the scenario. For example, if a flight that normally takes 1 hr is planned, but the pilots have been told to plan for 2 hr in the simulator, there is obviously additional time to be accounted for. The alternatives available under these conditions, with or without an ATC hold, will be apparent immediately to flight crews familiar with the geographical area. In addition, their behavior in the simulator can be influenced by their own speculative assumptions regarding the reasons for the inclusion of the additional time.

Pilots should fully understand the "game plan." Inadequate briefings have created problems in LOFT, and can create greater problems for LOS research. Unless contraindicated by the research objectives, some familiarization with the study objectives is desirable. The crew will certainly know that they are involved in some sort of research. Not only will a briefing of objectives help them bridge the gap between the real and the simulated world, but if they are left in the dark, at least some of them will try to deduce the desired behavior, and modify their normal performance. This

point is crucial and should be stressed when routine line behavior is desired. Even with an ideal prestudy briefing, it is difficult to avoid a certain amount of "Hawthorne Effect."

An unfortunate by-product of subject anonymity or confidentiality, which is a requirement for most research projects, is a diminished personal identification with the outcome of the simulator exercise. Fortunately, this may be a minor consideration. Pilot egos are strong. Under nearly all conditions, they will try to produce a professional performance. In the preflight briefing it is important to stress the point that the pilots were selected because they are professionals, and that the research is dependent upon their professional performance. The research studies that are probably most sensitive to this issue are those that can produce degraded performance (e.g., because of severe fatigue, or for any other reason).

It is critically important to point out that all supporting aspects of a regular line flight will be available. There are substantial differences among airlines, and the only reliable source regarding a specific airline's procedures is someone who knows that airline well.

Once the simulation becomes airborne, the same rules apply. Full company radio facilities should be available at all times. Pilots not flying should perform their functions exactly as they do on the line. Required operational report forms, including log book and emergency or irregularity forms, should be provided and used as is appropriate.

The crew should be asked to role-play exactly as if they were on a regular line flight. It should be stressed that if any events which are not a part of the scenario (including simulator malfunctions) occur, the flight crew will be informed immediately. If this is not done, there is always the possibility that the flight crew may mistake a scenario-induced problem for a simulator malfunction. It also can help save a research run in which a simulator malfunction does occur.

RUNNING THE SCENARIO

The schedule must allow plenty of time to get started because there can be many last minute details that require attention by the researcher or the flight crew. As in LOFT, there should be no interruptions of the scenario once an FMS has begun. There should be no observer or researcher interface with the flight crew other than in a simulated crew member exit from the cockpit, or in a simulated visit from a cabin crew member.

The only exception is the case where the scenario must be interrupted to change simulator configuration or collect data that can be gathered in no other way. In these cases, one has to create plausible events (see appendix B for a visual system changeover), or interrupt the scenario at natural break points. We emphasize, however, that scenario interruptions should be made only as a last resort and that the researcher

will have to take special precautions to avoid any "cascading" effects (appendix E).

Deviations from the Scenario

There is a high probability that there will be deviations from the scenario, and that the scenario operating team will have to cope with them. Deviations can come from straightforward operational decisions, such as the time when decisions are made to divert to an alternate or the amount of time a crew is willing to hold with a given amount of fuel. Many deviations are predictable, and should be included in the scenario as scripted alternatives. In some cases, subscenarios may be needed to get the flight back on the track.

Captains always have a final "emergency authority," which permits them to take any action which in their judgment is necessary to preserve the safety of flight. This can include actions such as landing below minimums, proceeding without or refusing an ATC clearance, and diversion to an unauthorized airport. This authority is not used often in real-world flight operations both because "emergencies" are rare and also because most pilots have an antipathy to writing reports of any sort -- especially reports which automatically trigger an official investigation.

The researcher should be aware of the captain's emergency authority and should know that it can and has been used in simulator exercises. Here traditionally, there has not been (but in research there should probably be) a requirement to complete an emergency report after landing and defend the action taken. Regardless of the research protocol, the area of "declared emergencies" (whether realistic or not) can be one of the most difficult areas in which to achieve the equivalent of real-world performance.

Unexpected Poor Performance

On rare occasions, an obviously poor performance, which can include classically poor judgment or even simply poor role-playing, can create a problem for the researcher. Failures do occur during the routine training and checking of experienced crew members in regular airline operations, and although the failure rate is very low, one has to be prepared for this possibility in LOS research.

The data secured in any instance of unexpectedly poor performance may or may not be useful. Although this is a judgment call for the researcher, the more important scenario issue is that such failures create a situation that must be handled with a great deal of tact during the rest of the simulator run, and during debriefing. Appropriate contingency plans should be made during scenario construction.

Simulator Crashes

Simulator crashes, including landing short during low-visibility approaches and overruns on short and slippery runways, can be in the same category as poor performance and need careful, reasoned consideration. Some researchers, believing the simulator should not be allowed to crash, will stop the simulator to prevent a crash, and then blame the incident on a simulator problem. Whether or not this procedure is desirable, it is not always possible, particularly in a low approach, aborted takeoff, or landing overrun situation. Fortunately, many incidents can be treated as minor, but controversy remains on the issue of whether or not to let the simulator crash.

One of the characteristics of LOFT is the lack of any intervention or interaction by the instructor or observer. LOFT flights are not interrupted for any reason, and continue to their completion up to and including realistically simulated crashes, if that would be the operational outcome from similar performance in the real world. FAA requirements for Phase II and Phase III training in simulators include "...the sound of a crash when the simulator is landed in excess of landing gear limitations."

Lauber and Foushee (1981) have cautioned that "an 'accident' should never be the inevitable outcome of a (LOFT) scenario, although it is always possible that one will occur." They also have noted the observation of airline training managers that "If an accident does occur during a LOFT session, it may provide the crew with a vivid learning experience." In the military, simulators are used for combat training and getting shot down or crashing is not an uncommon experience; however, military combat pilots know the risks and are prepared for them. The civilian pilot population does not have the same attitudes, values, or mission.

Except on rare occasions, a crash should never be the planned outcome of a research scenario. Unfortunately, an unwanted crash can occur in LOS research, as it can in LOFT. If it does happen, the simulated crash can create an additional problem for the researcher who is interested in creating as close to a real-world environment and reaction as is possible.

In the real world, "postaccident anxiety syndromes" have resulted from the acute situational anxiety which sometimes arises when a flight crew member survives an accident, and particularly one in which there were fatalities. The results can be severe. There are cases (see Popplow, 1984, "After the Fire-Ball") in which postcrash anxiety became so disabling that pilot careers were forced to be terminated despite psychiatric counseling and acceptable postcrash demonstrations of pilot proficiency.

We found only one reference to a potential psychological or psychiatric problem associated with simulator crashes (Lager, 1965), but the increase in simulator realism, strong ego involvement of professional pilots in their performance, and the LOS practice of not interfering even if the simulator

is about to crash, suggests that postcrash anxiety could arise in susceptible subjects. The possibility of a postcrash anxiety syndrome raises ethical and legal questions about the responsibilities of both researchers and research sponsors to protect the well being of human subjects.

Legal opinion is beyond the scope of this report or the competence of the authors. The researcher should review Federal Regulations on the use of human subjects for research, and obtain legal counsel. In general, the regulations require that all human subjects must be volunteers; that the risks must be defined and made known to them in advance; that volunteer subjects may withdraw from the experiment at any time; that adequate safeguards and facilities must be provided to protect the subjects; and that the research must be conducted so as to avoid all unnecessary physical or mental discomfort, suffering or injury.

The important point is that each researcher should be aware of this issue, and must decide whether or not a potential crash is to be allowed to continue to its conclusion. Each case will have to be decided on its merits with something very close to an instantaneous decision. If a crash is to be permitted, the subjects should be prepared for it during their indoctrination, and the organization performing the research should be prepared to handle a postcrash anxiety syndrome if the crash cannot be prevented. If crashes are to be diverted, then all possible conditions which might lead to a crash will have to be known, procedures and scripting developed to handle the problems, and the scenario team will have to be trained to recognize the situation quickly and execute the recovery procedures.

Simulator Problems

Simulator problems, including those induced by the research team or a simulator failure, are much more likely than a crash. One major reason for operator team training and simulator shakedown runs before the experiment is started is to minimize these kinds of occurrences. Simulator maintenance records should be reviewed to determine the most likely failures. The recovery procedures should be part of the scenario, which may have to be altered as unexpected problems surface during the shakedown runs. Providing for a greater number of trials than the absolute minimum required for the study is one way of coping with these issues after all methods of circumventing them have been exhausted during study preparation.

DEBRIEFING

Debriefing of the flight crew is an important part of LOS research. It is important to the crews who are understandably curious about their contribution and performance, and it is

important for the researcher because this is the optimum time to get reasonably uncontaminated subjective data from the study participants. Debriefing is the best time to discover the covert thought processes behind the operational decisions made.

Although debriefings can include structured or unstructured interviews, postflight questionnaires, video tapes of crew interactions, and so forth, they should start with an open-ended review of the flight by the flight crew itself. It is important to get their first impressions and overall reactions before specific research areas or audio or visual playbacks are discussed. Although it is impossible to avoid a certain amount of trying to "please the researcher" (who may be viewed by many of the participants as a prestigious authority figure), this tendency can be minimized if it is made clear to the flight crew that the researcher considers them the operational experts and wants and needs their expert opinion.

Researchers should remember always that the flight crew participants usually are, in fact, bona fide SMEs who have had an opportunity to view the simulator exercise from an important vantage point. The postflight debriefing is the ideal time to get crew reactions to the simulation and the scenario, and to explore the reasoning they used in reaching the decisions made during their flight. Audio and visual playbacks of the exercise are an effective method of providing "base points" and "reminders" for this part of the debriefing.

There are advantages in having a full crew debriefing, so that crew interactions can be observed and consensus can be achieved. In many cases each crew member will have a slightly different view of the events. On the other hand, there are also advantages to individual debriefing — the principal advantage being that the results will not be dominated by the strongest personality. Both practices have been used effectively.

A disadvantage with the isolated interviews is the additional research personnel that may be required to interview two or three crew members without forcing some pilots to simply wait their turn until a preceding interview has been completed. It is not easy to do this without adversely affecting the quality of the succeeding interviews. In some cases, it may be possible to mitigate this problem by having the pilot waiting to be interviewed fill out a postexperiment questionnaire on the scenario which has just been completed.

Finally, and the point is worth reemphasizing, airline pilots are professionals who have understandable sensitivity regarding their reputation. Positive aspects of their performance should be reinforced. They should be thanked for their contribution. It is particularly important not to infer to any crew member that they have performed poorly, or that they had more problems than others have had.

Poor or below average performance may need a rationalized explanation. "Simulator problems," or an allegedly "unrealistic scenario," sometimes can be used to help explain this very sensitive area. If possible, subjects should be

promised that the results of the study will be mailed to them when they are published. They should leave the facility with a positive reaction to the research project and a feeling that they have made a positive and professional contribution to research in aviation.

LEAD-TIME CONSIDERATIONS

An unfortunate fact, well known to those who have done LOS research, is that the lead time required to prepare for the study almost always will be underestimated because of the number of potentially critical considerations that are beyond the researcher's direct control. Preparation for the study can take much more time than it takes to run the experiment. The following list, and it is not a complete list, shows the kinds of items which should be considered. Each of them can take a substantial amount of time.

1. Development of the scenario and the training and coordination of the scenario development team. (See appendix G for an example of the time that can be required, even with highly experienced personnel.)
2. Interface with the employer and the pilot representing organization of the selected pilot population, if required.
3. Production of the scenario script.
4. Development and procurement of background environmental materials.
5. Procurement of manuals, forms, and trip paperwork.
6. Simulator scheduling.
7. Procurement and installation of any hardware or simulator software changes required.
8. Procurement and installation of data collecting devices.
9. Development and testing of data collection and evaluation materials.
10. Training of the experiment support team.
11. Scenario testing and revision (includes data collection).
12. Scheduling of pilots and outside support personnel.

The actual amount of time that should be allocated to accomplish each of the foregoing tasks can vary tremendously depending upon the study requirements, the facilities available, the make-up of the research team, and the familiarity of its members with line-oriented, full-mission behavioral research. There is a very high probability that the preparatory steps will take considerably more time than was initially allocated.

PREIMPLEMENTATION CHECKLIST FOR LOS RESEARCH

This chapter has discussed many of the practical issues involved in conducting LOS research, particularly scenario

and scripting requirements. The following is a checklist of items that should be considered in LOS research studies:

Conceptual Stage

1. Clarify the operational (or practical) problem.
2. Define the research objectives.
3. State the research question in researchable terms.
4. Determine the data needed.
5. Select methods that will obtain the data needed.
6. Determine the level of fidelity required. (Do you really need LOS?)

Development Stage

1. Determine scenario elements.
2. Procure scenario development team. Supplemental expertise may be needed in:
 - a. Operation to be simulated (type and location).
 - b. Aircraft characteristics.
 - c. Air traffic control.
 - d. Simulator characteristics.
 - e. Writing scenario scripts.
3. Develop and test data collection and evaluative materials.
4. Select research operating team. Individuals/skills needed:
 - a. Experiment coordinator ("wagon boss").
 - b. Pilot familiar with the aircraft and its operations.
 - c. Operational expert familiar with operation simulated and its ancillary services.
 - d. Data collectors.
 - e. Observer (researcher) in cab - may also have to operate simulator.
 - f. Air traffic controller.
 - g. Simulator operator.
 - h. Simulator technician.
5. Develop scenario: Select operational problems that can be expected to produce the desired behavior.
6. Write scenario script.
7. Administrative tasks:
 - a. Determine simulator availability.
 - b. Procure relevant software.
 - c. Procure and install necessary hardware.
 - d. Procure background environmental materials.
 - e. Procure and install data collecting devices.
 - f. Interface with employer and pilot representing organization if required.
8. Train research operating team.
9. Determine subject pilot requirements (qualifications and numbers).
10. Determine training (including differences training) and indoctrination requirements for pilot population selected.

11. Test and revise the scenario.
12. Develop preflight briefing material and protocol.
13. Develop postflight debriefing material and protocol.
14. Determine performance measurement requirements.
15. Determine data reduction and analysis procedures.
16. Develop data reduction and analysis software.
17. Perform sensitivity analysis of measures.
18. Schedule pilots.

A FINAL COMMENT

Past experience demonstrates that once a decision is made to conduct LOS research, few short-cuts can be taken. Line-oriented simulation research is both equipment and labor

intensive, and takes several months of calendar time. A complex LOS study easily could take more than a year from inception to completion. More often than not, the requirements of time and resources will be underestimated by all but the most seasoned LOS researchers. LOS is an exercise in details, all of which are important, and many of which can compromise the results of the study if they are not attended to with accuracy and precision.

However, good LOS research is well worth the efforts required. It is providing new insights into critical areas that many believed were not researchable, including many behaviorally related issues that have been called the "last frontier" in air transport safety. Conducting such research is a challenging task and there are substantial rewards for doing it well. The challenge to the researcher is limited only by his or her imagination.

CHAPTER 4

RESEARCH RECOMMENDATIONS

INTRODUCTION

The purpose of this chapter is to recommend research to improve full-mission LOS research, and not to address the broad issues of simulation and training. A thorough review of simulation issues from the perspective of the behavioral sciences was conducted by the National Research Council, Committee on Human Factors, Working Group on Simulation (Nones et al., 1985). A study of simulator training requirements and effectiveness was conducted by Semple et al. (1981). These reports address major issues in simulation and, especially in aircrew training, should be consulted for broad issues and research recommendations.

After examining the techniques of LOS research discussed in the previous chapters and in appendix C, it became evident that we need to know more about the use of LOS in behavioral research. Increasing the available knowledge can lead to increased confidence in conclusions drawn from LOS research, as well as to increased research productivity. The following key issues have not been answered well to date.

1. Full-mission research validity. Does LOS that is conducted in accordance with the guidelines presented in this report produce an environment in which the exhibited behavior is the same as would be exhibited in the real world?
2. Alternative forms of valid simulation. Can the behavior currently studied with LOS research be produced with abstract or part-task simulations? Stated another way: What are the research criteria for using LOS, and are there ways of obtaining valid behavioral measurements at less expense, with greater efficiency and control?
3. Optimization of full-mission research. Can LOS research methods be optimized? Are there alternate methods to achieve greater realism?
4. Human performance development. Can current analytic tools be improved, or new ones be created for assessing LOS features for a given application through the development of models of human behavior?
5. Subjective measures of fidelity. Can the elusive qualities of fidelity and validity in LOS research be measured by a simple, direct, and efficient method of subjective measures derived from expert personnel who serve as LOS research subjects?
6. Integration of research efforts. Can several of the above areas of research be studied together?

FULL-MISSION RESEARCH VALIDITY TESTING

Problem

In spite of extensive efforts to achieve high fidelity, the possibility remains that the behavior exhibited in the simulator is not the same as the behavior that would have occurred in actual flight. In fact, line pilots serving as subjects occasionally indicate that they would have behaved differently in the airplane. Although there is a suspicion that this sometimes may be a rationalization for poor performance, it might NOT be. Expert pilot comments may be indicating a simple truth and a simulation deficiency, whether or not that deficiency can be articulated clearly. Whatever the case, researchers should know the limits of their tools.

Even with optimum fidelity, a subject flight crew can never be expected to think they are about to fly a line flight while they are climbing the stairs to enter a research simulator. It is not clear at what moment (if ever) the crew members become so engrossed in the simulated problem that they have totally forgotten that they are in a simulator. Complete preoccupation with a challenging simulator task does not guarantee that all aspects of the preliminary steps, including the drive to the simulator location instead of to an airport, will stay obliterated in the pilot's memory. The behavioral implications, if any, of these questions are not known.

Approach

One ambitious approach to answer these questions is quite clear: Fly identical full-mission scenarios in both the aircraft and the simulator, collect a battery of measurements, and test the degree of correlation between the measures taken in the two environments. A major problem, aside from the enormous expense involved, is that of collecting in-flight and simulator measurements under identical scenarios.

Since it is not possible to specify in advance all the details of an actual flight to correspond with a scenario designed for the simulator, one must first measure, and record during flight, all basic parameters including communications, weather, and so forth, and then attempt to fly an identical scenario in the simulator. This process can be continued until a representative and sufficient sample of flights are recorded, providing a set of scenarios for duplication in the simulator. This would provide a paradigm similar to that of a backward transfer of training study (table 4-1).

There are a number of practical and methodological problems associated with such an undertaking. One is the difficulty of recording data during actual flights, and of achieving combinations of malfunctions, weather, traffic, and so forth that produce a wide range of behaviors. The task of data recording for a broad range of behavior could be impossible without using a special instrumented aircraft (or an elaborate video tape and manual data reduction effort). A large number of flights might be required to achieve data collection for a desired range of behaviors.

TABLE 4-1. - COMPARISONS TO VALIDATE FMS RESULTS

Scenario	Aircraft		Simulator	
	Crew 1*	Crew 1*	Crew 2	Crew N
1	X	X	X	X
2	X	X	X	X
3	X	X	X	X

*Same Crew

A methodological problem is created if the same crew is used for both flight and simulator because the simulator performance will be contaminated by the previous flight experience. For example, would the same decision making have occurred without knowing what had happened during the prior actual flight, or would different decision making occur because of "lessons learned" about what had happened previously?

Once the simulator scenario is established, data can be collected with a number of simulator crews. However, since people are unique, one would again expect a wide range of behavioral outcomes because of intersubject variability.

Expected Results

Data would be produced which should reflect on the validity of FMS research. Serendipitous information on how to improve FMS research also may be secured. The magnitude of the required effort is unquestionably large. This recommendation is made only because of the potential importance of the results; we may never know the validity of simulation-study results until such an effort is undertaken.

Alternative Approaches

Because the recommended effort is so potentially expensive and might never be done for that reason, research should be conducted to find ways to approximate such a study. As

an example, there are many aviation incidents every year, as well as accidents; these events are documented in data bases such as those maintained by the NASA Aviation Safety Reporting System (ASRS) and the National Transportation Safety Board (NTSB). These sources could be screened for types of events that have occurred several times, and those types of events could be placed in scenarios in simulators. If the response to the realworld events were replicated in a simulator with new crews (using similar equipment, operating procedures, crew experience, duty time, schedule constraints, and ATC and meteorological environments), there would be some evidence that performance in the simulator was representative of real-world performance for that type of event.

This alternative approach would require a substantial amount of work. The circumstances surrounding highly publicized incidents and accidents might be recognized by the crews, and their behavior might be altered by their prior knowledge. The unpublicized incidents (or published background data) might provide the most fertile source. One would need to enrich the database information with details which would be sufficient to construct a scenario event; a call-back to the crews involved (or a follow-up questionnaire) would be needed to derive such data. Also, incidents would be a better source of data because the flight crew members are still around to help recreate the scenario. Then a scenario might be constructed to include the circumstances which led up to the event in much the same way as LOFT scenarios, which are often derived from events that really happen in airline operations.

There are methodological issues that would have to be addressed, such as (but not limited to) how many occurrences of similar (to real-world) behavior would be needed to conclude that the behavior didn't occur by chance alone? Conversely, what conclusions would be reached if none of the simulator crews replicated the real-world performance?

It is not our purpose to fully develop such a method. We suggest that there might be some practical ways, short of the scientifically best way we have recommended, to demonstrate that performance in the simulator is representative of real-world performance for specific classes of events.

ALTERNATIVE FORMS OF VALID SIMULATION

Problem

There is an informed (but unproven) consensus that high-fidelity FMS techniques provide insurance against the possibility of conducting research on behavior that is substantially different from that which would occur in the real world. It is believed that this approach ensures that the human operator collects and processes information in the same way as in the real world. Its substantial advantage is

that it avoids the difficulty and almost prohibitively expensive task of performing complex behavioral research in the real world.

Unfortunately, research productivity can be quite low since high-fidelity, full-mission research is time-consuming and costly. It is also always possible that the full-mission approach will include a host of extraneous factors that may mask the results of interest. Finally, at some time the interaction of effects must be studied, not just results embedded in a typical situation. In this case, there may be a need for simple and truncated environments for comparison with full-mission results. In short, there may be types of studies which can be accomplished better and more efficiently using other research methods, including less than FMS. The problem is that researchers today are on uncertain ground in many cases and, understandably, elect the safer route — maximum fidelity in a full-mission context.

Approach

When using less than FMS, it is NOT necessary (although it might be desirable) to compare results to real-world in-flight performance, since performance in the high-fidelity, full-mission simulator can be used as a baseline or criterion. The a priori assumption is that the high-fidelity, full-mission simulator does produce the equivalent of real-world behavior. Therefore, the approach is to perform a high-fidelity, full-mission study, and then use a number of levels of simulation with the same purpose or subpurposes.

The final step would be to correlate the results to see if equally valid data was obtained using simpler devices. The degree of comparability, of course, will depend on the nature of the behavior involved. While control behavior may not be measurable with all abstract simulations, it is conceivable that all forms of decision behavior may be measurable using significantly lower levels of abstraction.

This approach is a variant of the approach used successfully by the Visual Technology Research Simulator (VTRS) behavioral research program at the Naval Training Equipment Center (cf. Lintern, Wightman, and Westra, 1984). In the VTRS program, transfer effects of reduced visual system, simulator, and motion system fidelity are estimated by a "quasi-transfer" study which measures the transfer of training within the simulator from the reduced simulation to the highest-fidelity simulation. The approach permits many factors to be screened in the simulator environment, and both transfer of training and performance effects can be measured.

It is recommended that a matrix of alternative simulations be used, as shown in table 4-2, with levels of abstraction as one dimension of the matrix, and part/full-mission simulation as the other dimension. A large number of levels of abstraction are possible, including high-fidelity simulation, simulation without visual and motion, general aviation simulator (e.g., GAT-1 class), microcomputer with CRT display,

keyboard, and joysticks (with approximate aircraft dynamics), microcomputer with CRT pages and keyboard (no dynamics), and pencil and paper. It is believed that full-mission and part-mission simulations are possible with each of these devices. The part-mission simulation would treat only key portions of the mission, within or out of the context of an entire mission.

The highly abstracted simulations require some amplifying description. The CRT/JOYSTICK simulation assumes a rough model of aircraft dynamics: a display of instruments; joysticks to control pitch, roll, and thrust; and a keyboard to control avionics, aircraft configuration (gear, flaps, and spoilers) and subsystems (electrical, hydraulic, pressurization, and so forth). An improvement of the popular "flight simulator" software for microcomputers is envisioned.

TABLE 4-2.—COMPARISONS OF ALTERNATIVE LEVELS OF SIMULATION

	Isolated segments	Connected segments	Full mission
Full-mission	X	X	X
Fixed/no vis	X	X	X
General aviation	X	X	X
CRT/joysticks	X	X	X
CRT/keyboard	X	X	X
Pencil/paper	X	X	X

Each member of the crew could have a microcomputer system or terminal, and a flight or flight segment could be flown by a single pilot or a crew using the devices and whatever ATC communications and ancillary services are needed. State-of-the-art microcomputers, networked where necessary, could provide all the capability needed for airports, radio navigation aids, scenario control, and performance measurement at a fraction of the cost of an FMS.

The envisioned CRT/KEYBOARD simulation would present a page of text and/or graphic information. The user would be present with various options, including decisions to take action or requests for more or specific information. Based on the user's action, another page of information and options would be presented. The experience would be analogous to reading the script for a play, but with the possibility of the play branching out in many directions depending on the actions of the reader.

Either booklet or interactive computer media could be used, with the user being directed to turn to a designated page with the booklet, or a new display being generated automatically with the use of a computer. The computer implementation would have the advantage of allowing complex algorithms for determining the next display, and also would allow automated measurement of user selections.

Full crew participation with multi-terminal configurations are possible with a computer implementation. The users could be "walked through" an entire flight, step by step, in the context of an FMS. For part-mission simulation, the users could have a similar treatment, but for only a portion of the mission beginning with a display of the initial conditions for that segment.

Expected Results

It is expected that this avenue of research would show that some forms of study can be conducted more economically with little disadvantage compared to high-fidelity, full-mission research. This could result in greater research productivity. On the other hand, if the results indicate the necessity for full-mission research, that level of research productivity must be accepted, and any extrapolation of laboratory research in those areas must be suspected.

OPTIMIZATION OF FULL-MISSION RESEARCH

Problem

Two additional problems can be addressed as corollary activities in any study of alternative forms of simulation. The first of these three problems, which are actually a set of problems, occurs because of the multitude of decisions required in the design of LOS, full-mission research. For example:

1. Is a day-night visual system required, or will a night-only visual system suffice?
2. Should there be multiple air traffic controllers (and different voices), or will one professional air traffic controller suffice?
3. Is anything really gained by beginning the mission with a long low-activity (i.e., normal) segment, or can one go directly into high-activity segments which will provide the primary research data? If so, should an initial portion of the high-activity segment be considered a warm-up?
4. Will broad classes of malfunctions provide essentially the same decision and crew coordination tasks?
5. Are there scenario-independent measures of behavior which would be valid and meaningful for research?

These kinds of questions raise the possibility of generalized scenarios and measures. The questions arise because, in cockpit resource management training, the specific system failures and scenario events are not as important as the fact that the scenario produces a complex problem for crews to solve. The flight crews have to exercise good resource management skills to solve the problem safely and efficiently. Traditional pilot-system performance measures of deviations

from a known profile capture only a small part of the important behavior.

The second problem occurs because of the extensive number of preexperimental missions which must be flown before a complete and refined scenario is developed. This is time consuming and expensive. The preexperimental testing may be as much as half the total effort. Attempting to shortcut preexperimental testing, which can be a very great temptation, is likely to result in many surprises during experimental testing. The result may be that much of the data collected has little or no meaning.

Approach

The first problem can be treated by experimentally comparing alternative forms of high-fidelity FMSs. For example, a scenario can be done by starting with a low-activity segment, and then repeated by beginning with the high-activity segment that is expected to provide the primary research data.

The second problem may be approached by using low-fidelity abstract forms of simulation (perhaps the booklet or computer form) to implement the detailed scenario for review by expert flight crews. For example, an entire scenario can be documented in booklet form, and then comments can be collected from SMEs. Revisions can be made to the booklet, and then more SME data can be collected. It is possible that refined scenarios could be quickly and inexpensively derived in this way.

The booklet approach can provide leads to the measurement issues, but it is unlikely to solve them. Behavioral measurement issues require continued effort and research in areas which are under investigation by NASA at this time (e.g., workload, communications, and performance measurement), and in human-performance model development.

Expected Results

More refined and efficient forms of FMS should result from this study. The study could lend credibility to FMS results and give insight into key scenario-design components.

HUMAN PERFORMANCE MODEL DEVELOPMENT

Problem

Full-mission research attacks more complex aspects of human performance than have been treated extensively in the past. Included are the domains of human performance termed supervisory control and cognitive processes. It is highly probable that a model reflecting a better

understanding of human behavior, even though approximate and incomplete, can be used to improve the design of FMSs.

For example, at the level of simple task analysis, a knowledge of the stimulus and response requirements for a given task can be used to determine if the simulation would permit behavior as it occurs in the real world. If the simulation does not provide the information or permit the response the model indicates is needed, then one may conclude that real-world behavior is not likely to occur. This is an example of the classical mixing together of empirical and theoretical attacks to form a synergistic relationship.

Approach

Two areas of model development which appear to promise important insights into the human performance involved in FMS are Supervisory Control (Sheridan, 1983) and the area of artificial intelligence called Expert Systems (Crowe et al., 1981; Obermayer et al., 1984). Both are especially important to the design of FMS, even when the research is directed toward other specific issues, such as the social interactions of crew members.

The supervisory control approach can structure and identify the multi-level control tasks of the crew. Analysis of cognitive processes in terms of models used for expert systems will allow extraction of specific rules used by crew members in making inferences and taking action. Expert systems have been developed for an array of sophisticated applications. They have considerable promise for a much needed representation of decision behavior.

Extraction of data for model development during FMSs, and then a subsequent testing of the models against human performance in other FMSs, is the recommended approach.

Expected Results

The short-term result expected is a specific framework for describing human behavior in FMS. This framework would focus attention on the behavior which the crew exhibits, the manner in which it is exhibited, and the resulting correspondence to real-world behavior. Improved understanding of this behavior will have many benefits, including improved analytical tools for the design of full-mission research studies.

SUBJECTIVE MEASURES OF FIDELITY

Problem

There is a need for a quick and easy test for satisfactory levels of simulation fidelity. Such information may be

obtained subjectively from expert users. The expert user, such as the experienced line pilot, is the possessor of detailed system-specific knowledge and experience — a level of knowledge and skill not easily acquired by a task analyst.

Ostensibly, the expert user can compare the FMS with prior experience and evaluate the simulator capabilities. A judgment can be elicited as to the estimated differences between simulator and flight behavior. An instrument is required which will do this in a valid and reliable way. No other approach is as tractable as subjective measurement with the expert user.

Approach

A structured questionnaire must be developed which can ask detailed questions about each of the features involved in the simulation across each segment of the mission. The fidelity of the simulated task can be compared with a subjective judgment of the level of fidelity required to perform the task, as it is done in the real world. The development and testing of such an instrument is substantial but can be accomplished.

For example, each question must be carefully phrased and tested for correct interpretation by the appropriate population. The effect of the order of questions and orientation of scales can be moderated through randomization (suggesting computer implementation). Data reduction and analysis software can be developed using algorithms for attitude measurement to improve interpretation of results. Thorough validity testing should be accomplished by comparing the subjective measurement with other corroborating data.

Expected Results

Given appropriate development, a reliable and valid instrument is expected to result. It is an instrument which can be used easily, and can be expected to achieve widespread use. Therefore, it is important that an appropriately designed and tested instrument be developed before an incompletely developed method becomes a defacto standard.

INTEGRATION OF RESEARCH EFFORTS

Although specific areas of research are discussed separately in the preceding sections, they need not be studied separately since studies can be designed to address two or more of these topics at the same time. For example, the validation and alternative levels of simulation problems can be merged; the LOS research data collection could be used for comparison with in-flight data as well as with data from abstract simulations.

The data collection for comparison between alternative levels of simulation can be combined with data collection for model development (with data collection for both purposes being derived from the same subjects). The subjective measurement development can take place with any of the other simulation studies. Furthermore, an experimental design can be developed which combines the study of alternative levels of simulation with alternative factors for optimizing LOS research.

CONCLUSION

A detailed plan of research depends upon the tradeoffs between available resources and competing goals. While such an attempt is outside the scope of this study, we believe that the research recommendations presented here can provide presently unavailable and needed knowledge regarding the use of FMS in applied behavioral research. We also believe that these recommendations can improve the utility of LOS as a research tool.

APPENDIX A

FIELD INTERVIEW NASA STUDY OF PREFLIGHT AND POSTFLIGHT OPERATIONAL PERFORMANCE IN SHORT-HAUL OPERATIONS

Dr. H. Clayton Foushee

RESEARCH GOAL

The purpose of this study was to examine the effects of duty-cycle exposure on an airline crew as a function of an actual line trip. Two groups of subjects were used: those who went directly to the simulator at the end of a 3-day trip (postduty) and those who performed on the simulator after 3 days at home (preduty). An observer noted critical events and rated performance during each simulated flight. Video tape recordings were made so that similar ratings could be performed by a panel of experts at a later time. Critical flight parameters were recorded onto floppy disks from the simulator computer and were time-synchronized with the video tape recordings.

GENERAL DESCRIPTION OF SCENARIO

It was desired to collect data in the context of a realistic flight rather than what may be judged as a contrived sequence of emergencies. It was desired to develop a scenario so there would be no question that the results could have happened during airline operations. Key decision points were designed into the scenario. Other conditions were defined so that generally the same flightpath should be selected by all crews.

Consistent with the availability of subjects and a simulator, a short-haul flight was planned to start at City A and end at City B. The weather was generally bad, and the aircraft was heavy with minimum legal fuel. Some equipment was inoperative at the start. At takeoff time the airport would go below landing minimums. At City B the crew would find that conditions were not suitable for a Category II landing and, if a landing was attempted, they would find that the actual ceilings were below decision height. The only alternate destination with acceptable weather conditions, given the fuel state, would be City C.

A "System A" hydraulic failure would be introduced while executing the missed approach procedure at City B. This failure requires manual actuation of the landing gear and electrical actuation of flaps, which is very, very slow. Also, the leading-edge flaps and gear cannot be raised once they are lowered, and a 15°-flaps approach is required because the

flaps cannot be raised fast enough to execute a missed approach. Additionally, there is reduced effectiveness of thrust reversing and the anti-skid systems. As a result of the malfunction, the crew must modify their normal approach to City C and may declare an emergency condition. The final destination was characterized by hazardous terrain and a short wet runway which had additional implications for the malfunction.

Key decisions designed into this scenario are: 1) requesting more fuel during flight planning; 2) requesting a take-off alternate, since immediate landing at the takeoff location is not possible; 3) determining that landing at City B is not suitable for Category II operation, because of the crosswind component; 4) determining that City C is the only suitable alternate; and 5) coping with the System A malfunction.

SPECIFIC SCENARIO CONSIDERATIONS

Subjects

Arrangements had to be made with both labor and management organizations to acquire subjects. This, together with the availability of a suitable simulator, narrowed subject selection to one airline. The scenario was therefore tailored to the requirements of that airline. Each subject was informed that involvement in the study would be anonymous and that data would be identified only by a code number. Otherwise, subjects could fear that they were being given a checkride which might influence their employment.

All of the subjects were volunteers. Because of flight crew scheduling realities and the nomadic behavior of pilots, it was often difficult to schedule full line crews for the simulator runs. Despite a high level of interest and very good cooperation from the pilots. It was considered essential to have a full line crew for each simulator exercise. Scheduling difficulties have prolonged the time required to complete this phase of the study.

Each crew was given a preflight briefing which stressed the importance of their participation. They were told that the study really depended on them. They were asked to role play -- to fly the simulator and make any decisions exactly as they would on a line trip. Full dispatch and all other pre-flight services would be available, and they would be

expected to complete all preflight papers in the same fashion they are filled out for line flights. Care was taken not to tell crews how long the exercise would take since this would indicate that a normal flight would not occur.

Observer/Ancillary Support Personnel

Individuals trained in weather, and ATC were used to make communications to the airline crew. A script was used to provide routine communications consistent with the scenario, as well as to simulate communications to other aircrews. Background communications tapes were piped into VHF 1 and 2 to simulate other traffic and ATIS information. Information was included in communications to other aircrews which might also be used in decision-making by the experimental crew. In addition to script communications, the support personnel had to provide any information which the crew might request. An observer was positioned in the simulator flight cabin, and while this might present a deviation from realism, this intrusion was considered necessary for data collection. The observer also provided functions for any other support personnel, such as passenger-cabin crew personnel.

Flight Planning

Actual company trip paperwork and preflight planning were performed as they would be for a normal flight. An experienced dispatcher was available. The first decision for the flight crew involved the flight plan. They had close to minimum weather, a heavy airplane with a lot of payload, and planned fuel that was legal but less than they were normally used to taking. They could add fuel but the dispatcher discouraged this because it would reduce their payload. There was a developing front along their route which was causing rain and generally low ceilings throughout the area. In addition, the copilot was able to start role playing by setting up a "cold" aircraft.

Weather

Care was taken to ensure that weather could have been typical of weather previously experienced in the scenario area for spring through fall. Experienced weathermen were consulted. It was desirable to channel the flight to City C so that experimental flights were controlled to a common flightpath; also, it was judged by weathermen to be typical of the area that City C could be "open" when the rest of the area was below landing minimums. Lightning flashes, turbulence and rain showers were used in the visual scene to alert the flight crew and corroborate with reports of deteriorating weather conditions.

Simulator Equipment

Some equipment was not available in the simulator, namely, radar, ACARS and the number 3 VHF. The experimental crews were briefed that this equipment was "inoperative." Modifications to the procedures were necessary as a result. The crew had to be briefed on alternatives to the number 3 VHF, since it was normally used for the acquisition of ATIS. The crew had to advise ATC that radar was inoperative. Simulator malfunctions had to be identified and fixed during the ground flight-planning period. Experienced flight crews were used during pretest to ensure that simulator procedures and equipment were company specific and consistent with those used by the experimental crews during normal line flights at the airline in question.

Everything about the flight was made as realistic as possible. It started with a pushback after the cargo and main cabin door lights indicated closed and the flight crew received an appropriate message from the cabin crew and clearance from the ground. The timing of the pushback with activation of simulator motion jars the simulator to realistically suggest a pushback with a tug.

Taxiing in the simulator presented a problem since the visual system did not have sufficient field-of-view to permit right turns. The scenario had to be designed so that only left turns were required for taxiing.

High-Low-Workload Periods

The scenario design produced a flight which started with a low-workload flight and ended with a high-workload flight segment. This permitted an analysis of behavior during both high- and low-workload conditions. However, the initial low-workload conditions, typical of most airline flights, provided a period for the crew to develop a realistic mind set and alleviate "simulator syndrome." It was the belief of the investigator, based on past observations, that crews are generally suspicious upon entering the simulator and abnormalities introduced early in a scenario tend to reinforce these suspicions. By letting the crew relax, the probability of their behavior being realistic when a problem does occur is increased. Such judgments about realism as well as judgments about the level of workload, were made as the result of extensive pretesting flights.

Malfunctions

It was desired to include a malfunction which required high-level decision-making but did not pose a serious hazard to flight safety. It was desired to end the flight in a safe and satisfactory manner. The chosen malfunction required time-consuming manual deployment of gear and electrical extension of flaps. The malfunction was introduced while

executing a missed approach at City B and there was a problem associated with the time of the malfunction. They were cleared for an approach but the controller gave them a revised wind that exceeded their crosswind limitations. If the crew, while flying the missed approach, were slow in raising gear and flaps, it was possible that gear and leading-edge flaps would be irretrievably locked in the down positions. This would complicate the fuel situation and make it very close on fuel to City C. In addition, the malfunction had additional implications for landing on the short wet runway at City C.

Ethics

It was considered unethical to expose flight crews to experiences which could be psychologically damaging. No flight was allowed to deteriorate to a crash which in the real world would have killed hundreds of people. Consequently, the investigators had to be very careful in the design of the scenario so that such things wouldn't happen. There is no guarantee, however, that conditions might not deteriorate badly at the end of the flight at City C. While it might be necessary in some cases to terminate the simulation early, this did not occur.

Debriefing

The crews were instructed not to discuss any aspect of the simulator experience with any other flight crews so that

future subjects would not be contaminated with such knowledge.

SCENARIO GENERATION PROCEDURES

The scenario was developed in three iterative steps: First, scenarios from LOFT were collected from airlines using this technique, and were reviewed for application to this study. The initial scenario adopted was based on these considerations, and was reviewed by various SMEs from the selected airline.

Second, supporting materials and support personnel duties were developed. This included development of communication tapes, procedures, performance assessment techniques and video recording methods.

Finally, the scenario was flown during extensive pretest flights. Fifteen full scenarios were flown and revisions were made as anomalies were noted. The testing, together with review by SMEs, is the key to developing a realistic scenario. There is a tremendous amount of detail which can affect realism, and many pretest flights are required to achieve a scenario which will not contain unrealistic elements that can affect the behavior of operational crews.

The amount of work preparing for the study is approximately equal to the amount of work expended during the remainder of the study.

APPENDIX B

FIELD INTERVIEW NASA FULL-MISSION SIMULATION STUDY OF CREW COORDINATION AND DECISION MAKING

Miles R. Murphy

RESEARCH GOAL

The primary objective of this study (Murphy et al., 1984; Murphy and Awe, 1985) was to develop methods of quantifying crew coordination and decision-making factors and their relationships to flight task performance. A secondary objective was to develop information about crew process and performance for application in the development of resource management training programs. Of special interest was obtaining information on how errors evolve in the cockpit, particularly errors involving interpersonal factors.

Relationships between several crew and systems performance measures and some personal and crew process variables were explored in this study. Personal variable categories include personality and background variables, such as age and experience. The primary emphasis, however, was on crew process, or interpersonal interaction. Constructs, or variable classes of major concern, were: 1) command hierarchy, 2) command style, 3) interpersonal communications, 4) crew coordination, 5) resources management, and 6) group decision making.

GENERAL DESCRIPTION OF SCENARIO

The scenario represented a flight from Tucson (TUS) to Los Angeles (LAX) via Phoenix (PHX) with a forced diversion to an alternate upon reaching LAX. The crew's enactment of the scenario began with a Captain's briefing in the simulated operations room at TUS and ended upon deplaning at the selected alternate — either Palmdale (PMD) or Ontario (ONT).

The scenario was designed to evoke a series of decisions about where to proceed following a missed approach at LAX caused by a nose gear "not-down-and-locked" indication. This situation was exacerbated by having occurred at a time when the Los Angeles basin, including Ontario, was experiencing low and deteriorating ceilings and visibilities caused by coastal fog. Ontario, located inland from Los Angeles, was lagging Los Angeles in this deterioration. And, just over a mountain range, out of the basin, Palmdale had clear weather with good visibility. Upon going through a complete gear check procedure taking several minutes, the crews would

discover that the gear was down and locked, and they could therefore assume that the panel light indication was faulty.

Within this scenario the most critical dimensions of the decision process involves when to proceed from the LAX area to an alternate airport, and the choice of the alternate. Related subsidiary choices involves whether to do a complete gear check in the LAX area, whether to make a second landing attempt at LAX (ceilings and runway visual range (RVR) degrade to legal minimums at LAX during this choice "window" and will go below minimums if and when the aircraft crosses the outer marker), whether to raise the gear for fuel conservation while flying to the alternate, and whether to declare an emergency for either the gear problem or a minimum fuel problem.

SPECIFIC SCENARIO CONSIDERATIONS

Simulator

A Boeing 720B flight training simulator, a later version of the Boeing 707, was used in the study. This simulator was an FAA-approved visual simulator with a model-board scene and had 3° of freedom in motion: pitch, roll, and heave. The simulator was operated by the Airline Training Institute (ATI), San Carlos, California.

Subjects

The subject crew members were paid volunteers. Their experience represented a wide range in reference to airline of origin and recency, or currency, on B-707 line operations. Some were current on the B-707. Many had recent B-707 line experience and were now flying other jet aircraft in line operations. Some were retired from the line. This diversity in experience was considered important as an aid in evaluating the sensitivity of the various performance measures. Thus crew composition ranged from one in which all members were retired from the line to one currently flying the B-707 as an intact crew.

This diversity dictated that special-differences training be administered to review knowledge specific to the simulator operations that might be different from current or previous

line operations. All crew members received 6 hr of classroom differences training and 4-8 hr of simulator differences training. The number of hours of simulator differences training that a crew member received was based on recency. Subjects were formed into crews prior to this simulator training and were instructed in coordinated procedures during this training. Some baseline flight task performance data were obtained for each of the two pilot crew members at the completion of the simulator training.

Experimenter Team

A current, professional air traffic controller was used in the simulation. The controller also participated with another member of the experimental team (the pilot advisor) in simulating conversations with other aircraft, thus providing background conversations on the ATC network. Two observers seated in the back of the simulator rated crew performance and did experimenter tasks required within the simulator.

The air traffic controller and other experimenters were located in a control room adjacent to the simulator. Monitors available there were an X-Y plotter showing the aircraft path, a visual scene display, audio speakers, and video screens showing views of the crew members and cockpit. The total experimenter team consisted of nine persons - including two persons for simulator operation and reconfiguration. (Reconfiguration of navigation receivers, airport parameters, and so forth, was required periodically to simulate the complete flight route.)

The experimenter team had to be prepared to deal with unexpected events, such as the accidental movement of a cockpit lever leading the crew to believe that a malfunction (not a part of the experiment) had occurred. The air traffic

controller had to be prepared to deal with any type of information request that the crew might generate. Timing of communications had to coincide with specific events, rather than at designated times, complicating execution of the scenario by the experimenter team. The team, in communicating to the subject crew, had to be extraordinarily careful not to "lead" the crew in decision making, or to add distracting reminders of unrealism. The ATC and background conversations were scripted, although occasional contingency intervention was required. Fuel available at the initiation of the approach to LAX was standardized by clearing the aircraft from an enroute hold when the fuel level reached 14,000 lb.

Training for the experimenter team consisted of briefing sessions and rehearsal during the two partial, and one complete, "shakedown" simulator runs. In retrospect, some additional training and rehearsal may have been appropriate considering the criticality of effective coordination within the team during data simulator runs. A further consideration here was that some projected team members, who had participated in scenario development and its adaptation to the simulator, required replacement shortly before the start of simulator runs (caused by experiment delays and changes in those team members' situations).

In summary, preparing for a large complex study like this one can involve more time than planned - particularly when contract administration issues are involved. Unanticipated loss of experimenter team members can be a problem. Contractual time limitations (e.g., for the simulator use) and budget limitations can constitute a pressure to do less "shakedown" and experimenter team training than may be desirable. Some considerations in training time requirements may be the extent to which each team member is both research and operations oriented, and whether multi-organizations are represented within the team.

APPENDIX C

APPLIED BEHAVIORAL RESEARCH

INTRODUCTION

This appendix discusses the nature of applied behavioral research in general terms. It provides an overview of the factors that influence the character and conduct of applied behavioral research in a contemporary aviation context.

The main topics discussed are the conditions that affect the research process and the scientific and practical goals the researcher is striving to attain. The research process has two principal stages: planning and execution. Both stages involve continual resolution of conflicts between the ideal and the real — between ultimate goals and means. The discussion will deal not only with the fundamental issues and factors involved in research planning and execution, but also with the compromises and tradeoffs that the behavioral researcher is bound to encounter.

RESPONSIBILITIES OF THE APPLIED BEHAVIORAL RESEARCHER

The responsibilities of the researcher are threefold: 1) to satisfy the scientific requirements; 2) to satisfy the practical requirements; and 3) to manage the research project. The scientific and practical requirements often conflict, and many factors affect the way the study can be performed.

It is not uncommon for a researcher to strive for scientifically credible results without articulating the scientific criteria to be met or the rules to be followed. Both of these are classic errors. Similarly, it is not uncommon for the researcher, particularly if one is new to the applied behavioral research field, to lose sight of the practical needs and expectations of the customer. It makes little difference whether the customer is someone in a different department of the same organization, higher level management, an outside agency, the operational community that uses the equipment or the system of concern; or more amorphously, another discipline such as designers or engineers.

The following paragraphs discuss several frequently underestimated real-world considerations as well as the general scientific and practical requirements that should shape the character of an applied research project. A basic assumption is made that any individual responsible for research on human behavior that is aimed at answering a practical (operational) question understands the accepted scientific method. A further assumption is that he or she also understands the

problem domain — or has enough sense to go out and learn about it before undertaking the research.

Scientific Requirements

General goals and methods— The object of scientific inquiry is to describe, explain and predict natural phenomena. Science is an activity involving observation, theory and practice. Its goal is reliable knowledge (Morris, 1955). While rules change, there are common conventions within the broad field of empirical science that make a study acceptable. And beyond that there are special articulations of these conventions that conform to the “established viewpoint” within each paradigm of normal science (Kuhn, 1970).

The conventional requirements for credible empirical research and results are straightforward and disarmingly simple. These requirements include an operational definition of variables (i.e., observable or reducible to observable events), repeatability of findings, and exclusion of alternative explanations. Desirable characteristics include quantitative relationships, parsimony (i.e., simplicity of explanation), and generalization of results.

Behavioral goals and methods— Although behavioral research methods have some similarity to those used in the physical and biological sciences, they are tailored to the special problems of behavioral research. Human behavior always is influenced by multiple factors regardless of the apparent simplicity of the task of interest. As Uttall (1981) pointed out, too many things, both external and internal, affect behavior to expect to find that a single, simple stimulus has a prominent, predictable influence on an overt act.

Humans vary in their knowledge, abilities, experience, and attitudes. Humans are also adaptable and changeable. That is, they can learn quickly to adjust to situational demands, and their performance will change with time as a result of experience, motivation, fatigue, and other factors. The individuality, complexity, and changeability of human behavior all have important methodological implications that must be taken into account when performing behavioral research.

Behavior usually is thought of as an overt manifestation of internal processes. The goal of the scientific study of behavior is to understand those processes — not merely to understand the overt act. The same act can be the result of many different internal factors. Behavioral research methods are necessarily complex and rigorous because assurance of the validity of the results comes more from the soundness of the testing methods than from the apparent behavioral outcomes.

Scientific skills of the behavioral researcher— The behavioral researcher applies general and specific scientific, methodological skills to the study of behavior. Because behavioral processes are not directly observable, the researcher's first requirement is to reduce behavioral problem statements into operational terms (Carnap, 1955). Here the required skill is largely a product of understanding scientific method. This is most important in applied behavioral research because the problem questions, which are raised by the people who have to deal with them in the real world, are almost never stated in a directly testable form (Cody, 1984).

Knowledge of experimental design and statistical methods are equally important where measurements are made on properties in a sample of the members of a population. This is true whether the measured properties change with time or whether the sample population consists of humans or non-living objects that are not identical in all relevant respects. Experimental design is a specialized skill that encompasses selection of testing methods, variables, and subjects in a manner that permits generalization of results and the application of inferential statistics.

Use of descriptive and inferential statistical methods is required to express data values in terms of the population sampled in a manner that encourages confidence in the accuracy of these values. Within behavioral science there is an implicit expectation that the relationship found or effects produced will be probabilistic in nature. There is an equally explicit understanding that if action is suggested or required, an inferential statistical test should demonstrate that the results or conclusions have less than a 5% probability of having occurred by chance alone.

The behavioral researcher also needs the ability to make behavioral measurements. Too often researchers have been trained only in the rudimentary aspects of measurement — i.e., those that are chiefly concerned with controlling the conditions of measurement, avoiding contaminating effects, and restricting the range of options available to the test subject. Development of meaningful measures is invariably a significant part of the research problem if the behaviors of interest and the testing situation are even modestly complex.

Practical Requirements

The principal practical requirement is to secure information to answer the problem question. Frequently, the customer has expectations about how the research should be done, and the form and characteristics of the information that will be derived from it. Almost always there is an expectation that the research will focus on the specific problem of concern and be performed in a context that is a realistic representation of the actual or expected operational environment.

For example, the problem may be to determine the degree to which certain head-up display (HUD) symbology will

obscure a pilot's view of a runway during landing approaches. The researcher might see this as an instance of a more generic issue, such as the detectability of certain classes of objects as a function of the percent and distribution of lines on a HUD, and as one which could be better investigated in the laboratory.

The customer, on the other hand, is likely to have minimal interest in the general problem. He or she can insist that only the HUD configurations of interest at the moment be used, and that the study be performed in a flight simulator using experienced pilots. Additional customer expectations may be that any effects found will be large and obvious; unambiguous; that any systematic effects will be in the form of simple, direct relationships, without a long inferential chain of explanation; and that the results or conclusions will be expressed in terms of the customers' jargon or expertise. These are realistic expectations. The customer wants results that can be understood and reasons to have confidence that the results will be valid in the real world.

If the customer's background is in engineering, which is common in organizations dealing with equipment and systems, proof or validation of hardware and software concepts is usually tested at each stage of building a prototype system. Care is exercised to perform thorough tests. The outcomes are usually clear — something either works or it does not. When measurement is involved, a tangible property is measured.

Such clear outcomes are seldom the case in behavioral research. The customer may not fully understand the required methodology, and may be assured of the validity of the results only if operational people are the subjects and the testing situation resembles the real world. Quite simply, the customer (or the ultimate consumer) frequently requires the behavioral research to have face validity. Face validity is easily understood. It is the customer's principal source of confidence in the credibility of the behavioral research.

Real-World Considerations

Real-world constraints and the resources available are crucial considerations. Time and money are both enabling and restricting factors. Each is a sensitive area, and together they largely dictate the scope of the study. Once the researcher makes an initial estimate regarding the time and money required, the estimate becomes a commitment, and an almost inevitable constraint. It need not be said that the time and money required are often underestimated.

Resources— Facilities, equipment, and a variety of human resources with appropriate skills will be needed to perform the research. Human resources include the research team, support personnel, and subjects. The interests and expertise of the researcher team influence the way the problem will be framed and approached. Supporting personnel include those people who provide services necessary to

the planning and conduct of the research project but are not part of the research team.

The number and types of supporting personnel will vary with the scope of the project and the research vehicle. Equipment technicians, computer programmers, and usually a data analyst will be required in virtually all cases. For simulation-based studies of moderate to large scale, one or more SMEs – usually experienced operational personnel – will be necessary. Simulation support personnel include simulator facility managers and technicians, specialized computer programmers for real-time control, and scenario generation and visual database programmers if a computer-generated image system is used.

Other human resources needed are administrative support personnel and, in many cases, a representative of the customer, and finally, the test subjects are a particularly important human resource. Their availability and characteristics affect the scope of the study as well as the research vehicle that may be used.

Facilities and equipment determine the choice of research vehicle. The reliability of the test and data acquisition equipment should be known or determined in advance. Data acquisition capabilities affect the choice of measurement methods.

Specialized knowledge, which is a critical intellectual resource, affects both the research plan and its implementation. At least three kinds of specialized knowledge are required for behavioral research:

1. Basic knowledge of behavioral processes and the operational context in which they are used. This knowledge largely determines the selection of conditions and variables to be used to answer the research question.

2. Knowledge of experimental design. First, to properly use the design selected, avoid or control unwanted effects, and secure data which is amenable to statistical analysis; and second, to be familiar with a breadth of different experimental designs so that the design used is appropriate for the complexity of the research project and its economic constraints.

3. Knowledge of performance measurement. The development of such measures can be a significant part of the research effort. The goal of measurement is to produce information that is meaningful at the operational level. To do this, one needs criteria for determining in operational terms how a performance is good or bad. For many types of applied research, such criteria are unknown. The development of meaningful performance criteria can be a significant part of the research effort.

Constraints— Real-world constraints also affect the research project. They must be taken into account because behavioral research takes place in organizational and social contexts that impose their own requirements or restrictions. These constraints can come from three sources: the customer; the subjects; and the management of the researcher's organization.

The constraints imposed by the customer are the customer's requirements and expectations. These have already

been discussed. Customer constraints can have serious consequences because they directly affect the complexity of the research vehicle, the subject population, and the form of information required for a "successful" research project. Unless the customer can be accommodated, or persuaded to accept a modification of perceived and strongly held requirements, there may be no research project.

A basic subject constraint is imposed by the ethical requirement to respect the privacy, dignity, and self-esteem of an individual. While in some cases it may be desirable to produce research situations that explore these areas, as well as others where organizational or social pressures influence performance, these are exactly the circumstances that are considered to violate the individual's personal rights.

Subjects can also impose additional constraints when they are expert operators of a system. Understandably, they can be expected to have a critical view of the research representation of their familiar working environment. When the research vehicle includes an FMS, expert operators can be expected to be sensitive to the details of the simulation, even if certain details have no relevance to the research goal. There is a considerable risk in not accommodating the expectations of expert operators, because any reservations they may have about the simulation can affect their motivation and performance.

The third set of constraints on behavioral research can arise from the requirements of the researcher's organization. Demonstrations or tours for visitors or the press often will be required for high visibility projects or facilities. They should be planned for in the research schedule. Priorities among projects may influence accessibility to research facilities. Contractual obligations may require completion of work steps by a certain time or in a particular order, even though the research schedule might be better served by a different schedule. Coordination with other organizational units may be necessary, particularly if the behavioral research is a part of a larger project.

PURPOSES OF APPLIED BEHAVIORAL RESEARCH

Research is a means to acquire knowledge for an end purpose. It may involve efforts as simple as the searching of a single document or as elaborate as a series of experiments over several years involving construction of facilities and equipment and the training of a staff. Here we are concerned only with a particular kind of research – scientific investigations to better understand the behavior of humans interacting with machines.

Research is not automatically required to solve every problem. It is needed only when knowledge from other sources is suspect or insufficient. Therefore, the first step when confronted with any problem is to formulate a precise description of the information needed to solve it. The second

step is to gather existing information and, if additional information on the subject is needed, to consider the alternatives for obtaining it. These steps are obvious and important. They are stressed here because if these preliminary steps are not done carefully, the end result can be inappropriate or unnecessary research.

The lack of confidence that existing information is satisfactory or directly applicable to the problem is responsible for a great deal of research. For example, many applied studies have been performed on the effect of platform motion on performance and learning in a flight simulator (see Waag, 1981 for a review). Yet, it is still impossible to marshal strong support for either a yes or no answer to the question of whether platform motion is needed for a particular purpose.

Would performing another study be useful? The answer might be "yes," if the argument is that conditions in previous studies were different from those currently of interest, and therefore the results do not apply to the present problem. The answer might be "no," if one looks closer and finds that the difficulty is not differences in the conditions, but in obtaining definitive performance results. The point is that early effort in the problem solving process can have significant downstream effects on research decisions.

The nature of the problem determines the purpose of the research and, in general, there are three main purposes of applied behavioral research: hypothesis testing, problem exploration, and evaluation. Each is discussed in the following subsections.

Hypothesis Testing

Scientific research is commonly thought to consist mainly of hypothesis testing, in which predictions are made about the relationships between variables. The object of the research is to confirm or disprove these predictions. The form of the predictions may be correlational or stated as cause and effect. The principal features of hypothesis testing are 1) the formulation of the hypothesis at some general level of description in terms of the variables, and 2) being able to demonstrate that the predicted relationship holds over a range of specific situations, and is not accounted for by other factors common to those particular instances.

Problem Exploration

Problem exploration is research to discover and isolate factors that contribute to a problem. For example, if the problem is the occurrence of several incidents of aircraft descending below glideslope during night approaches under visual meteorological conditions, the purpose of the research is to discover the factors that lead to these occurrences. The principal objectives of problem exploration are to precisely

describe the conditions and states of variables present in several instances of the situation of concern, and to describe the relationships in time of the associated actions or sequence of events.

A main distinguishing feature of problem exploration is that it is atheoretic. This is as it should be, because the researcher should have no firmly preconceived ideas about the causal or correlational relationships among the variables. This sometimes is difficult in practice, for in many cases there will be at least a notion of broad classes of variables or conditions that are thought to be important. More detailed data are likely to be collected on these factors than on others. The researcher should be sensitive to the danger of missing or underestimating the importance of other factors.

No attempt should be made to control the operational situation in problem exploration. The primary objective is to discover the particular factors that are present or absent when the event of interest does or does not occur.

Evaluation

Research for evaluation is also atheoretic. Its primary purpose is to answer a question or solve a particular problem. Evaluative research has several forms, but usually involves comparisons of systems (or elements in a system) in an attempt to discover performance differences. Comparative studies often are done to find out if using one type of equipment, configuration, or procedure results in better performance than another.

Evaluation research also may involve comparison of performance against a predetermined criterion. This is usually done for validation purposes. For example, a study might be done to confirm that a simulator visual system can support the performance of all normal and emergency visually guided flight tasks.

A variant of evaluative research is to obtain baseline performance data, either for an existing system which does not have adequate baseline data, or for a new system. If a new system is being evaluated, the comparison will be made as a prediction, based on the development and implementation of new or different equipment, configurations, or procedures.

APPLICATION CONTEXT

The nature of systems change with advances in technology. At present, the aviation system is undergoing major changes in the underlying architecture of its control system as well as the appearance and function of its control and display mechanisms. Behavioral research topics are determined by the role of the human in these new or modified systems and by the practical problems that arise during their

development and employment. In the next sections, we will briefly review the nature of older and newer systems, the changing role of the human in the newer systems, and the implications these have for applied behavioral research.

Older Systems

An older system, including an aircraft of previous generations, is an assemblage of relatively discrete subsystems, each with its own set of displays and controls. Each control and display is dedicated to a function and has a fixed relationship to its subsystem. A control mechanism has a simple, direct effect on the controlled element. Similarly, a display instrument presents elementary data about some aspect of the subsystem.

In older systems, the human operator is responsible for closing the control loop between the displays and controls as well as the coordinated operation of the subsystems to achieve the general system goal. The number of people needed to operate an older system is directly proportional to the number of monitor and control functions required. For example, in older commercial aircraft which operated over water or over large, desolate geographical areas, five crew members were required (pilot, copilot, flight engineer, navigator and radio operator).

Human operators of older systems (including aircraft) fulfill their roles by monitoring and integrating information from several discrete sources and making continuous or very frequent control actuations. The operators' tasks, and hence their training, are basically manual control, or procedural in orientation. Manual control tasks require the operators to learn the system dynamics well enough to control the system with the precision and stability required. Aircraft configuration and subsystem operations (e.g., electrical, hydraulic, and avionics systems) are an example of tasks that are essentially procedural. That is, operation in normal and abnormal modes is largely a matter of following a set of rules of the form, "if x then do y ."

In older complex systems, tasks often are distributed among several operators. They frequently require a manager who also may have an active role as an operator, such as the captain of an aircraft, who receives information from the other personnel, directs and supervises their activities, makes all major decisions, and is directly responsible for the operation of the aircraft. Depending on the particular kind of system, operators (other than the manager) may not always be aware of the system's general status, can have a limited span of authority, and may have their activities confined to specific tasks.

A human operator will never be far removed from monitoring and control of the lowest functional level of an older system. Intermediate levels of control between the lowest level and the system manager also will be filled by humans.

This has advantages and disadvantages for responding to out-of-tolerance or emergency conditions.

A principal advantage is that humans are available to detect and respond to abnormal events within their span of authority or at least communicate status information to the manager. Fault identification is relatively easy since the machine parts of the system are distributed at a single level. Problems at intermediate levels of control are usually human problems and easy to distinguish from the machine level. Verbal reports to the manager can be succinct and descriptive of the exact fault.

Some serious disadvantages of these older systems are that an operator may not notice the symptoms of an impending failure, be able to react quickly enough, or be able to take a required action out of his or her local span of authority. An event affecting several systems, co-located but separated in control, may require action at a higher level of control. In such cases the intermediate or general manager will have to integrate the individual reports, diagnose the common basis of the problem, decide on a course of action, issue orders, and follow up to be certain that the corrective action was taken. It is not always easy to do this in a timely and effective fashion.

Newer Systems

The trend of new developments in manned systems is twofold, 1) functional integration of subsystems, and 2) implementation of automation at multiple levels with a high-order, computer-based controller overseeing the full range of system functions. The role of the human is to exercise overall control of the system. Emphasizing the role of the human, systems of this type are called supervisory control systems (Sheridan and Johanssen, 1976). The general control structure of such systems involves four functional levels. A human operator at the top, a human interface system, a semi-autonomous task-interactive system, and the task or subsystem function at the bottom level (Sheridan et al., 1983).

The physical architecture may involve several computers and differ among various types of supervisory control systems such as power plants, chemical process plants, command and control systems, aircraft, and undersea vehicles. A supervisory control system may have several human operators, but the general role of each is essentially the same. Multiple operators may be present for the sake of redundancy, or because the system is a hybrid of the old and the new control structure with a single general manager.

Characteristics of a supervisory control system include autonomous operation with automatic control being exercised at several levels, filtering and processing of information presented to the supervisor, and indirect but high-order control that is goal- or effect-oriented. That is, supervisor com-

mands are interpreted into sets of specific, coordinated control inputs to subsystems at the task level. The mechanisms for information presentation and control inputs are multipurpose displays and controls such as CRTs, programmable-legend and function switches, touch panels, keyboards, and tracking devices.

Automation of the details of subsystem function control frees humans to extend their range of authority over the entire system, or at least large segments of it. The tasks of the operator as system supervisor are to set initial conditions, monitor overall function, make adjustments, and intervene as necessary. The supervisor attends to achieving major system goals instead of the implementation mechanisms. The role of the human supervisor is largely to provide intelligence and deal with the unusual or unanticipated. Intervening in the event of an out-of-tolerance or emergency condition is the supervisor's primary responsibility.

Contrast of Older and Newer Systems

Older and newer systems differ in the types of tasks which humans must perform and in the behavioral processes that are called upon to perform these tasks. Older systems require assimilation of discrete data on low-level functions of the system and discrete control of these functions. Older systems are physical-control oriented. What the operator must know and do are largely manual control and procedures which are determined by the physical properties of the subsystems and the characteristics of the display instruments and control mechanism.

Newer systems are information-processing oriented. Supervisors are several levels removed from the subsystem level of functions. Supervisors see abstracted, summarized, and selected data of major process functions and issue general, goal-oriented orders. Supervisor tasks are to interpret information, make decisions, and solve problems when they arise. There is less requirement for procedural knowledge and actions, and more requirement for rational thinking and judgment.

Implications for Applied Behavioral Research

Creating a workable supervisory control system is more than a simple matter of adding automation. Although automation frequently changes the supervisor's workload, it does not necessarily reduce it. In addition, to the degree that automation performs functions beyond human capabilities of speed, precision, or complexity, the supervisor cannot simply take over if a failure or unanticipated event happens. A different kind of coping strategy (than a take-over) is necessary. Moreover, if automation is controlling a high-level coordination of general system functions, its modus operandi may be so different from what a human would do that the supervisor

cannot follow what is going on, and therefore could lose the ability to anticipate or plan ahead.

There is a large number of difficult philosophical and technical problems which need to be solved to achieve practical supervisory control systems. For example, at the philosophical level, there are questions about the amount of authority and responsibility to be given to the computer-control system. At the technical level, there are questions of how to afford the supervisor access to information and allow control at the subsystem level if it becomes necessary, and how to portray this information and effect control. These kinds of questions, as well as a host of others, will determine the nature of applied behavioral research in the future.

Behavioral issues in older systems focus on what people can do, their capabilities and limitations. Tasks in older systems usually can be divided into discrete units involving a few sources of information of an elementary kind, and one or a few control mechanisms. The behavioral concerns center on how to design the displays and controls to minimize errors and maximize the speed and accuracy of the operator's performance.

Implicit in the legitimate interest in what people can do are the notions that the person has few or no acceptable options to do something different. Either system goals or equipment constraints limit these choices. Either the operators will simply try to do the best they can, or there is a criterion of what is to be done. These notions imply that there is a rule for the behavior.

Because the operator of an older system has a reasonably immediate control involvement with the subsystems, and the characteristics of the subsystems are essentially immutable (i.e., they have fixed mechanical properties or are hard-wired), the behavioral issues associated with man-machine interaction in an older system are superficial to some degree. Concern focuses on the operator interface with the display instrument and the actuator mechanism.

Operator performance quality in older systems is governed by these devices and the ability to use them in the prescribed manner. It has little or nothing to do with the general system goal or external interactions of one subsystem with another. The behaviors of interest, therefore, are mainly those needed to interpret an instrument (vision and perception), operate a control (motor-control speed, accuracy, and bandwidth), the linking process of memory, and the enabling process of learning. In effect, the predominant behavioral domain of interest in man-machine integration in older systems is stimulus-response relationships.

Performance of concurrent tasks and integration of information and its interpretation in terms of a general goal have always been a concern. Sources of information that frequently need to be integrated are placed in close proximity or incorporated in a single instrument (e.g., in an aircraft engine performance gauges are grouped together, and the pitch and roll indicators are combined in the attitude instrument). Applied behavioral researchers have investigated issues

of divided attention, vigilance, and compatibility of displays and controls in an effort to improve the human factors aspects of system operation.

Providing new capabilities, reducing personnel requirements, and reducing human error are incentives for improving a system. Solutions to human-related problems in systems have been directed toward automation. This has been approached largely on a piecemeal basis rather than as a part of an integrated development of the entire system.

The piecemeal approach has had adverse consequences for operators, particularly the crew of aircraft (Wiener and Curry, 1980; National Research Council, 1982; Sheridan and Hart, 1984). Automation does not necessarily relieve the operator of workload, but simply may shift the burden from physical activity to mental activity. The increased use of automation, however, is a trend toward fully realized supervisory control systems. The issues are no longer whether or not to automate, but how to integrate automation to have the human effectively act as a system supervisor.

The relevant applied behavioral issues extend much further into the interaction of the human with the system. Human supervisors are removed from the subsystem level, and have a broad span of authority and responsibility. Automation of routine manipulative- and continuous-control tasks, as well as those for which there are fixed procedures or rules, leaves the human to perform decision making and problem solving tasks. If more than one human is involved in the system, the supervisor's task will be to manage the human as well as the machine resources.

The supervisor will operate at the system goal level and, in effect, will be interacting with several levels of system function. Since routine problems are handled automatically, the difficult problems, those for which there are no predetermined rules, must be handled by the supervisor. This will require a greater in-depth knowledge of the system functions by the supervisor than is now necessary for the human operator. Predominant tasks of the supervisor will be knowledge-based rather than rule-based or skill-based (Rasmussen, 1983).

The behavioral factors of most concern will be information seeking, assimilation, integration and interpretation, decision making, problem solving, and resource management. The man-machine interface problem will involve integrating deeper levels of the system with higher-order cognitive processes. The research challenge will be to discover a means to achieve cognitive compatibility between an automated controller and the human supervisor to allow the supervisor to monitor the status of the system in terms of functional purpose, and to convey goal-oriented commands to the system. As Singleton (1976) has written, "The whole point of man as a supervisor as opposed to an operator is that he needs to be able to make intelligent responses which in turn implies that he reacts in terms of concepts and not in terms of stimulus-response units."

In general, applied behavioral research issues in the newer systems will shift in emphasis from what *operators can* do to what *supervisors will* do — or, for existing systems, what they *do* do.

In contrast to older systems, supervisors may have several options available to deal with a problem or achieve an end. The criteria for their behavior are not as clear cut, except in terms of the outcome. Also, the critical aspects of their behavior, information processing, and decision making will not be as accessible for measurement as a control manipulation. None of these changes will make it easier for the applied behavioral researcher.

In addition to the shift in the behavioral processes of interest, there is a trend to expand investigation of the scope of variables that influence behavior beyond the immediate situational or equipment characteristics. Perrow (1983) has pointed out that the organizational context (i.e., the social structure of the work environment, including management attitudes, peer pressure, and personal goals) influences the decisions people make in the operation of a system. As an example, Woods (see appendix E) stated that nuclear power plant operators take many more positive actions in a training simulator than they do in actual operations. He attributes this difference to the reward and penalty structure operating in the real work environment that biases operators to not take any action if there is uncertainty about the need or outcome.

RESEARCH VEHICLES

A research vehicle is simply the facility used to support the research. Research vehicles can be classified into three categories: the real world, simulation of the real world with varying degrees of complexity, and the laboratory. Each has advantages and limitations.

Real World

Except for research on new systems, the real world is the beginning and the end stage for research. The real world is appropriate for exploratory studies to determine the factors associated with a practical problem, and it is the final proving ground for evaluation of new developments studied in simulation or in the laboratory.

From the viewpoint of the customer interested in practical applications, it is the ideal vehicle for conducting a study. It is the only context where all relevant factors operate. Organizational goals and requirements, stress from real hazards, peer pressure, self-esteem, and long periods of routine activity are among the most important conditions that influence behavior. They occur naturally in the real world,

and are very difficult to create in other research environments. The real world is the best choice for behavioral research on what people will do in contrast to what they can do.

The real world, however, does have several disadvantages for research purposes. There rarely is an opportunity to control all conditions; rare events are almost impossible to study; hazardous situations cannot be duplicated; experienced subjects are necessary for a system of any complexity; and data collection is difficult and can be very expensive, particularly if systems such as an aircraft are required.

Simulation

Simulation includes a broad category of research vehicles. It ranges from comprehensive representation of the operational equipment and environment in support of full-mission performance to limited simulations involving a single item of equipment or a short task. The advantage of simulation is combining real-world hardware, most environmental conditions, and task demands with the ability to control events and conditions. Rare and hazardous events can be introduced, any required data can be collected, new equipment or procedures can be incorporated with the old, and usually data can be collected in a shorter period of time and more economically than in the real world.

Limited simulation is particularly useful for the testing of concepts or theory derived from basic laboratory research, or adopted from contexts unrelated to the one of interest. It also is well suited for investigation of problems that have been isolated by exploratory research in the real world or in FMS.

The disadvantages of simulators as research vehicles are generally proportional to their complexity. The disadvantages include the effort and difficulty required for setting up and controlling the simulation process, and the latitude of the alternative courses of action available to the test subjects. This latitude tends to make each run unique in some respects and, therefore, can create data analysis problems.

Either experienced or trained subjects will be required for simulation-based research that involves anything more than the most simple task. The subject training problem becomes difficult and takes on another dimension if the research addresses new equipment or procedures and requires the performance of complex tasks approaching a full-mission context (Sheridan and Hennessy, 1984). By definition, there are no experienced subjects for new developments; training of naive people can require days to accomplish, and can be very costly. If personnel experienced in an existing system with some similarity to the new configuration are used as subjects, they also will require a certain amount of training. With experienced subjects, there is great concern that old habits or preferences will influence their behavior, and there may be a negative transfer of learning.

Measurement is a problem in complex simulation studies. Although the type and amount of data that can be collected are large, considerable effort is required for the development of measures. Complex simulation is used for behavioral research only when high-order, complex, or subtle behavior is of interest and requires detailed examination. For example, the effects of fatigue on performance, decision making in unexpected or emergency situations, and supervisory control of automated systems are typical applied behavioral studies that could require complex FMS as a research vehicle. Because of the nature of these problems, performance must be measured in terms other than simple, overt acts or psychomotor control if meaningful aspects of behavior are to be captured.

Laboratory

The laboratory is best for initial test or elaboration of theoretical ideas about basic behavioral processes, preliminary research on equipment characteristics such as display coding and formatting, control methods and, to some degree, measurement methods. Its main advantage is the ability to isolate and study a particular behavioral process. Subject training needs usually are minor, and no special experience is required. The set-up of equipment is short and easy compared to configuring a simulator, and data can be collected quickly. The disadvantage of laboratory research for applied behavioral studies is its remoteness from the realistic context from which most practical problems arise. It is more suitable for study of generic issues than for specific application problems.

Comment

The levels of research vehicle described above are really points on a continuum. Even the real world and simulation overlap when some aspects of both are present. For example, the Total In-Flight Simulator (TIFS) at the Naval Air Development Center is an aircraft with two cockpits, one for research on control and displays, and a customary one for safety of flight. Moreover, the research cockpit can operate as a ground-based simulator. The military air combat maneuvering ranges are another example. Here weapons delivery and their effects are simulated, but the aircraft and air defenses operate in the real world.

Limited simulation and laboratory settings also blend together. New systems frequently consist of electronic displays (usually CRTs) and controls linked by computer to the actual hardware. In a sense, simulator technology has been incorporated into real systems. Since CRTs and computers have become principal tools in the behavioral laboratory, many features of real systems and simulators can be created there.

FIDELITY AND VALIDITY OF RESEARCH SIMULATORS

The required characteristics and features of the research vehicle are prominent issues when planning applied behavioral research. If a simulator is the research vehicle of choice, these issues will be considered in terms of fidelity requirements. In this section fidelity and the closely related concept validity are discussed as they relate to the determination of the requirements for research simulators.

The two principal factors that should determine the choice of research vehicle for an applied behavioral study are:

1. The type of research required by the problem (i.e., exploratory, hypothesis testing, or one of the varieties of evaluation), and
2. Knowledge of the factors that influence the behavioral processes of interest.

The type of research dictates the specificity of representation that the research vehicle must have. Knowledge of the factors which influence the behavioral processes determine how comprehensively the research vehicle must represent an operational system and its associated situational conditions. Together, these features are commonly thought of as the fidelity of the research vehicle.

Fidelity

Definitions— Fidelity is a confusing term. Much of the confusion is due to the tendency to talk in qualitative generalities about fidelity as if it is a single-dimensional characteristic and independent of a specific simulation and application. Although there is a lack of consensus on a definition of fidelity of simulation (Hays, 1981), it has generally been discussed from two major viewpoints.

The first viewpoint treats fidelity as a physical characteristic of simulator equipment. For example, Huff and Nagel (1975) define physical fidelity as the objectively measurable correspondence between the operational system and the simulator equipment in form and function.

The second viewpoint treats fidelity in terms of behavioral effects such as psychological fidelity or perceptual fidelity (Matheney, 1975), cognitive fidelity (Spears, 1983), or behavioral outcome (Jones et al., 1985). As stated by the National Research Council's Working Group on Simulation (Jones et al., 1985), "Certain difficulties in simulator design can be avoided if fidelity is defined in terms of potential effectiveness for a planned use rather than in terms of physical correspondence."

A similar view was expressed by Semple et al. (1981): "Aircrew Training Device fidelity is the degree to which cue and response capabilities in a simulator allow for learning and practice of specific tasks so that what is learned in the device

will enhance performance of these tasks in the operational environment." Obermayer (1964), Semple et al. (1981), and Hays (1981), as well as many others, have pointed out that fidelity is a multifaceted concept.

One facet of fidelity is abstraction. A simulator may consist of real-world equipment driven by a computer, devices that externally appear to be real equipment but are internally very different or nonfunctional, or devices that have no external resemblance to the actual system but correspond functionally in terms of a mathematical model.

Often it is assumed that departures from fidelity by abstraction are detrimental to effective use of simulation; however, abstractions are not always detrimental, particularly in training. There are several cases, such as cockpit procedures trainers in aviation, that portray some instruments with pictures because this level of abstraction is sufficient to fulfill the intended training purpose. STEAMER (see National Research Council (Jones et al., 1985) for a description) is a simulated propulsion plant of a Navy frigate represented schematically on a dynamic color graphics display and driven by a computer model of the plant. The simulation functionally corresponds to the real system, but is in no other way a physical representation of the propulsion plant.

Another facet of fidelity is accuracy. Real-world characteristics may be included at various degrees of precision. In some instances, reduced precision can be an advantage for some purposes, such as training. For example, there are cases where a full engineering mathematical model of actual flight dynamics produced a perceptually unacceptable control task, but the dynamics produced by simplified or adjusted models were perceived to respond more like the aircraft (NATO AGARD, 1980).

In these cases, the mathematical model had to be simplified by removing terms from the equation or reducing gains on some terms (by as much as 60%) to create handling qualities which were acceptable to pilots. "Tweaking," or modifying, of simulators to fly "like the real aircraft" is common, and is suspected to be needed because all the visual and motion cues of the real world are not fully or faithfully represented.

A third facet is completeness; some real equipment, environmental conditions, or agents may not be represented in a simulation because they do not affect the behavior of interest. For example, a study by Brown et al. (1958) used a centrifuge as a motion platform for a flight simulator and concluded that a fixed-base simulator would be as good for prediction of the simple tracking task performance they studied.

Fidelity in the choice of a research vehicle— It often is assumed that high fidelity is never a disadvantage, but there are research as well as practical reasons not always to strive for maximum fidelity. In general, high fidelity implies a comprehensive representation of the real world. The problem for the researcher is that in real-world situations, many factors acting in variable ways can influence behavior. In a great

many of them, there are a number of optional ways for an individual to behave.

Two basic principles of behavioral research are to maintain control of the research situation and to account for the factors which influenced the observed behavior. High fidelity simulation complicates the task of maintaining control. It provides an opportunity for unknown, extraneous factors to influence behavior, and gives subjects an opportunity to choose behavioral alternatives that may be beyond the research scope of interest. These effects show up as variability in the data and reduce the sensitivity of the performance measures as well as the reliability of their values.

From a practical viewpoint, high fidelity representations of real-world situations incur costs that are proportional to the comprehensiveness and complexity of the research project. The direct costs are for support personnel required, installation and maintenance of the equipment, and for the computational power necessary for real-time control of the situation. Indirect costs accrue from the time and effort needed to plan and execute the study, subject recruitment and training, the number and duration of the test trials, and the data collection and analysis procedures.

Often high fidelity representations can be undesirable because of the specific purpose of the research. If the purpose is to discover what people can do instead of what they will do, the research may require the elimination of features that can only degrade performance. For example, Fitts et al. (1958) commented about the ATC studies they conducted at Ohio State University: "One of the major tenets that we have followed is that human capabilities should first be determined under optimal system conditions (e.g., with 'idealized' displays and reliable information) and then be determined for suboptimal or degraded systems. Only data obtained under idealized conditions permit an estimate to be made of the upper limits of system performance that could result from future improvements in the machine aspect of the man-machine system."

In other cases, it may be desirable to depart from high fidelity by enhancing or augmenting real-world features for specific purposes. For example, computer-generated flight imagery may include supplementary cues, such as a highway in the sky (Lintern, 1980), to increase the precision or consistency of performance leading up to an event that is the subject of research interest.

The extent that fidelity is an issue of concern depends on the general character of the research vehicle. Fidelity is less of an issue for laboratory research than for simulation-based research. Laboratory research is primarily used as a research vehicle when there is no great concern that the behavior be identical to what would occur in a particular operational context. The purpose is frequently to test general behavioral principles or to determine the effects of a limited number of factors on a specific behavioral process, regardless of whether other factors will also have an effect. The conditions created in the laboratory can be abstract, limited, or both, and the

points of correspondence with an operational system often are irrelevant.

Simulation usually is chosen as the research vehicle when complexes of behavioral processes are of concern, or when the factors influencing the behavior are not well known and it is desired to elicit behavior equivalent to what would occur in the operational context. Uncertainty about what factors affect behavior is the primary reason to strive for high fidelity of simulation.

Early in research planning, the researcher should think about each potential factor that affects behavior, and determine the characteristics of the simulator necessary to produce the desired effects. There are several good reasons for this exercise:

1. The exercise will make explicit what is known and not known about factors influencing the behavior of interest.
2. The exercise will focus its attention to specific characteristics of the simulation for a particular purpose, instead of the vague notion of general fidelity. Producing a simulation that is comprehensive, precise, and a concrete (i.e., not abstract) representation of an operational situation is costly. To ignore the details of the simulation requirements is to ignore major cost factors.
3. The exercise will establish a set of criteria for the researcher to later assess how the final configuration of the simulator meets the research need.

Once the project is under way, it is easy for the researcher to rationalize why features of the simulation thought to be important at the outset of the project (but which could not be included) become unimportant when a practical impediment is encountered. It is better to document initially what the needs are thought to be, so as not to become self-deluded later. If a characteristic thought to be necessary cannot be attained, the researcher should think hard about its potential consequences, and of the alternatives ranging from restructuring the research plan to explicitly noting the potential implications for the validity and interpretation of the data.

Practical, as well as scientific, considerations will determine the characteristics of a simulator used for research. Simulators are expensive devices; they are seldom constructed for a special project and disassembled when it is done. There may be several simulators with differing features available for use and, apart from slight modifications, the researcher's options may be restricted to selecting a particular device with a relatively fixed set of characteristics. Typically, the simulators that are available determine what can be used.

It is worth mentioning at this point that although a simulator is a tool to support research, use of the tool rather than the research need can easily become the motivating force. There is a strong temptation to let the availability of a device dictate what research is performed. If a large simulator is available, researchers often will find a justification for its use. This is another reason the researcher should formally

establish the characteristics of the simulator required to support the research project. It helps to keep the simulator cart behind the research horse.

However, it oversimplifies the issue to say that a simulation should have all, but no more than, those characteristics that directly affect the behavior being examined. This is an ideal goal, but one that rarely can be achieved. For reasons which have been discussed, the customer frequently will require characteristics of the simulation that are stated vaguely in terms of fidelity. Although researchers frequently will acquiesce to this requirement, they should always evaluate the implications for control of the situation, and the variability of performance that might result because of the presence of extraneous factors.

Experienced operational personnel used as test subjects will also have expectations about the fidelity of the simulation. This can be both good and bad. Experienced operators, in voicing concerns about the fidelity of simulation, may identify factors that affect the behavior of interest in ways that the researcher had not conceived. This is obviously good.

It is bad, however, if the subjects develop a negative attitude because their expectations of fidelity are not met. This will inevitably affect their behavior. Fortunately, there are other ways to deal with this problem if it is impractical to meet the subjects' expectations of fidelity. The alternatives are to educate the subjects about the reasons for the simulation characteristics, motivate them through instruction, and train them in the simulator configuration used.

In short, the concept of simulation fidelity, although intuitively compelling and in widespread use, is difficult to quantify. It is equally difficult to determine the level of fidelity necessary for a specific application.

Validity

Including all factors believed to influence the behavior of interest is no guarantee that the behavior in the simulated situation will be identical to that which would be exhibited in the real world. It always is desirable to empirically determine that equivalent behavior does occur in the real world and in simulated contexts. This involves the validity of the simulation. It should be established whenever possible.

Definition of validity— Validity is defined as the statistical correlation between two sets of measures collected under differing circumstances. Depending on when and where the two sets are collected, differing types of validity can be defined (McCoy, 1963). The principal concern is predictive validity, the degree to which measurements made in simulation correlate with the same measurements made in the real world. If the two sets of measures agree, the simulation is considered valid for the conditions under which the measures were made.

Establishing the validity of a simulator is recognized as an essential requirement. Recently, a committee of the National Research Council was asked to assess means for improving the value of the Computer Aided Operations Research Facility (CAORF) as a research tool. CAORF is essentially a ship bridge simulator used for maritime research (National Research Council, 1983). The committee concluded, "The single greatest deficiency of CAORF is the *lack of validation* for its uses. Specifically, *CAORF's mathematical ship models and data on training and other human performance characteristics need to be compared to actual ship behavior and human performance in the real world* (underlined in the original). Validated models and studies would be CAORF's single contribution to maritime research and development, and need to be given top priority."

Verification of validity— The concept of validity is well defined, and verification is considered essential for simulators used for research intended to elicit real-world behavior. However, the testing which is implied rarely can be done easily, and often cannot be done.

Frequently, research is done in the simulator because it would be too dangerous or too expensive to do in the real world, or because the real system does not exist. Under these circumstances, validity testing is either extremely difficult or impossible. Practical constraints also minimize opportunities to perform validation studies. Applied behavioral research that is part of a system development program usually has a restrictive schedule and budget. There is rarely time or money to collect data "twice" to test validity. Consequently, few formal behavioral tests of simulator validity have been performed.

In the absence of direct measurement of behavior in real and simulated situations to verify the validity of a simulator, efforts often are made to establish validity indirectly by verifying the physical fidelity of the simulator. The rationale is that if the simulator is a comprehensive, concrete, precise, and accurate representation of an operational system, then the behavior produced should be equivalent to the behavior produced in the real world.

This point was made by the National Research Council, Working Group on Simulation (Jones et al., 1985): "...pilots now may be certified in a simulator that meets rigorous fidelity standards established by the Federal Aviation Administration. Such a simulator must faithfully duplicate physical and functional characteristics of an aircraft as well as the conditions of flight. Similarly, in engineering design, where critical and expensive design decisions may be based on performance in a simulator, high fidelity is the best insurance for obtaining valid performance data. For these applications, some lesser degree of fidelity may also produce valid performance data but it is usually not worth the cost or the risk to make the determinations experimentally."

Matheney (1978), in a discussion of the need for behavioral fidelity of simulators used for research purposes, introduced the concept of performance equivalence. Research of

the human pilot in vehicle control has advanced to the point where good models can be determined for the pilot control function (a process called system identification), allowing mathematical analysis and prediction for the overall control system. He concludes that we may be able to establish the behavioral fidelity of a simulator for the control aspects of a task through application of system identification procedures.

Caro (1977) proposed that flight simulator fidelity be assessed using a backward transfer paradigm in which experienced pilots are tested on their ability to fly the simulator. Significant deficiencies in performance would then be taken as indicators of simulator fidelity deficiencies. He states, "While backward transfer should not be the sole justification for simulator procurement, one would be hesitant to use a simulator which could not be operated by competent pilots."

Applied behavioral research is intended to gain new knowledge about human performance to answer a practical problem question. The "bottom line" is that without specific validation studies, or without specific theoretical knowledge about what factors influence the behavior of interest and how, the researcher is forced toward high fidelity of simulation to ensure that valid data will result, and to achieve acceptance by users of the data outside of the research community.

There is a lesson for research to be gained from experience with training simulators. It is evident, when considering transfer of training and training effectiveness, that much depends on how the training device is used. Although the characteristics of the training device (i.e., fidelity) can enable good training, it is up to the instructor and automated training features to produce a high transfer of training. Similarly, high fidelity may be an enabling factor for a research simulator, but the manner of use is a prime determinant of the validity of the data. More is involved than the physical and functional characteristics of the simulator.

This is especially true for full-mission aviation research which requires simulation of many real-world features in addition to those of the vehicle cockpit, visual system, motion platform, and equations of motion. Considerations of the fidelity of simulation must include the total environment, e.g., ATC, ground facilities, cabin crew, weather, and other aircraft.

Moreover, consideration must be given to establishing the framework of regulations, procedures, and preplanning which can have a significant effect on human performance. If behavior that skilled people exhibit in the real world is to be elicited, then the conditions that would exist in the real world must be established, including familiar missions and procedures. If the research is concerned with crew coordination and decision making, it must be recognized that the extended team includes ground team members in addition to those in the cockpit, along with all of the briefing and planning preflight activities.

CONCLUSION

All of the factors discussed in this appendix are part of the process of determining how a study is to be conducted and how the research vehicle is to be used. If the study is to be done in a real-world context, these factors should be addressed in the processes of study planning and experimental and scenario design. The experimental design, procedures, and scenario emerge from an iterative consideration of scientific and practical goals of the research, the available resources, and the constraints which are imposed. In this discussion, considerable attention was given to the issues of simulator fidelity and use of the term "validity" because an understanding of these concepts is central to the planning and execution of applied behavioral research studies.

APPENDIX D

FIELD INTERVIEW MARITIME BEHAVIORAL RESEARCH

Dr. Thomas J. Hammell
Dr. Myriam W. Smith
Eclectech Associates Division
Ship Analytics
North Stonington Professional Center
North Stonington, CT 06359

Tom Hammell (TH) is the vice-president for research and development of the Eclectech Division of Ship Analytics. His research using bridge simulators, including the Computer Aided Operations Research Facility (CAORF), primarily has been concerned with the fidelity and curriculum requirements necessary to use these devices for the training and licensing of mariners. Myriam Smith (MS) is a research psychologist in the same division who has been using their in-house bridge simulator for studies sponsored by the Coast Guard to evaluate a variety of aids to navigation. The following are their comments on the process of scenario construction, fidelity requirements, and related topics.

SCENARIO CONSTRUCTION

TH began by discussing the process of scenario construction and several factors which influence the character of scenarios for simulator-based experimentation. Ideally, constructing scenarios should be a formal deterministic process. However, at some stage, there will be unknowns. By adopting a set procedure, at least the points at which unknowns occur will be identifiable. He agreed that the principal factors that influence the scenario design include:

1. The program objective (the research question).
2. The simulator characteristics and availability.
3. The participants' characteristics and availability.
4. The characteristics of the real-world context.
5. The performance measurement methods.
6. Ensurance of performance differences among conditions.
7. User acceptance (making the scenario credible).
8. Known or previously determined variables that affect behavior.

A great deal of time is necessary to develop the scenario. First, the experimenter develops a skeleton scenario which takes into account most of the factors listed above. Next, the scripting of the scenario requires working with an SME to ensure that the conditions are credible to the participants

and to achieve the approximate level of difficulty desired. A run-through of the scenario(s) is essential. No matter how well the script is planned, unanticipated effects will become apparent such as conflicts between scheduled events, inappropriate timing of events, errors in the programming and unexpected or undesirable actions by the participants.

He said that the experimental objectives and performance measurement considerations, i.e., what can we measure, are probably the two most obvious and most important factors. The number of independent variables that can be tested or must be tested also affect the character of the scenario. In most experiments it would be desirable to test more variables than it is possible to access given the practical constraints of time, money and subject availability. An attempt is made to determine if the scenario can be arranged to squeeze in one or more extra variables without incurring more work, i.e., requiring the development of additional performance measures, or excessively lengthening or otherwise complicating the conduct of the experimental runs. The number of dependent variables that can be measured also influences the scenario design.

Other factors that influence the nature of the scenario are the creative-cognitive processes (CCP) of the experimenter which come into play after the formal, analytical part of the scenario development is completed. The formal analysis provides you with the basic elements that go into the experiment, including the experimental goals and the knowledge of the maritime environment you have or have gained from experienced mariners. The CCP comes in bringing the two together, choosing the particular embodiments and arranging their time, spatial, and contingent relationships.

Regardless of the number of factors involved, the scenario development process is iterative, not serial, in nature. MS said that the process can be diagrammed as a converging spiral path on polar-coordinate graph paper where the spokes on the graph are the various factors and the center of the graph is the goal of complete scenario specification. Thus, in the scenario-design process, you iteratively consider the factors involved and the trade-offs among them but at the same time you are converging toward the design goal, i.e., developing greater specificity as you go along. MS found this apt

analogy in a book on ship design (Taggart, Robert, ed., *Ship Design and Construction*, New York, SNAME, 1980).

MS said she had not really thought of creating scenarios as anything other than determining a matrix of variable and constant conditions for an experiment. If this is done in sufficient detail the choices that are left are inconsequential to the purpose of the study. Normally when a full-mission experiment is planned properly, the SMEs who are asked to review the planned scenario will usually suggest changes that do not affect the experimental purpose, or significantly affect planned circumstances and events. However, it does sometimes occur that an SME will make an important suggestion that does influence some major aspect of the experiment and does require a significant change in the plan. TH agreed, saying that it is relatively easy to come up with several detailed scenarios that fit the experimental design.

REALISM OF SIMULATION AND SCENARIOS

How artificial or realistic you make a scenario depends on several factors: the attitudes of the participants, the need for control, how performance will be measured, and the purpose of the study. Realism in maritime simulation is a real concern. Mariners used for experiments tend to be very operationally oriented; they do not have much tolerance for situations that are not realistic. On the other hand, the experimenter would generally prefer to create an artificial situation, albeit with complexity similar to a real harbor or coastal area, because it affords greater opportunity for control of the experimental situation and creating just the right conditions, a combination of currents, geography, ship traffic, weather and events to elicit the behaviors of interest to answer the experimental questions. However, it is difficult to convince the mariners that it is the same as duplicating an actual harbor or coastal area. The attitudes of the mariners are very different when a real location is simulated. They take the exercise much more seriously. This helps with their acceptance of the simulation but brings on the problem of the mariners pointing out every discrepancy between the simulation and the real-world setting. As MS said, "they tell you what's wrong with all the little irrelevant details."

Another reason for using realistic simulation and scenarios is that the customer who sponsors the work, e.g., the Coast Guard, usually expects a great deal of face validity. In one case, a study of the rules of the road, the face validity requirement would have meant a 4-hr scenario. Moreover, the mariners and the Coast Guard, for the studies it sponsors, want geographical realism, i.e., using an actual setting rather than an artificially constructed one. (The counterparts in the aviation community, e.g., the airlines, the FAA, and the airframe manufacturers, to say nothing of the engineering side of NASA, almost always have had the same attitude, and it is a real-world concern that should not be ignored.)

SUBJECTS

Experienced mariners do not like to be tested. They have no performance-based test initially and no recertification requirement. So it is hard to get them to participate in an experiment that is essentially a test. The strategy used by TH is to present the experiment as a "training" scenario and ask the mariner to run through it so they can discuss its training merits when it is over.

PERFORMANCE MEASUREMENT

In regard to performance measurement, TH said that you prefer to rely on automatically recorded data, but these measures tend to be very microscopic relative to the kinds of questions you are trying to answer, i.e., quality of decisions and ship handling. Invariably, you must include observational data by SMEs. TH felt it was important to obtain analysts' (psychologists') observations as well because they can look at performance in a more generalized way. They can view particular actions as examples of classes of behaviors, e.g., classes of errors such as commissions or omissions.

MS pointed out that if the nonobservational data support your hypothesis or are consistent and unequivocal, you have no incentive to closely examine or analyze the observational data. However, the nonobservational data are rarely so revealing or clean in MOS experiments that you do not need to resort to the observational data. Also, regardless of the type of data, you always collect as much as you can because you never know for certain what to expect. Consequently, you record all the data that are feasible to collect. This does not refer just to automatically recorded data. Video taping is almost essential for any complex simulation experiment. If nothing else, you often need it to make sense of data from other sources when something unanticipated or unusual happens.

REDUCED SIMULATION VS. FMS

Using a reduced simulation situation is useful because it economically focuses on the particular behavior of concern. Reduced fidelity or limited simulation is good for examining in detail a phenomenon that occurred in an FMS exercise. However, it often requires greater effort on the part of the experimenter to do all that is possible to ensure the performance of the participants will be valid. When limited simulation is used to explore preliminary concepts, it can also be very economical, particularly if many iterations are expected to be necessary to evaluate several alternatives or sharpen a particular concept. In most cases where practical decisions

must be made, the results obtained in limited simulation will require validation in a high fidelity FMS.

USING SMEs

One problem with using SMEs is the uncertainty of the attitudes they are likely to adopt. That is, some mariners are very concrete in their thinking and are greatly influenced by any details of the simulation that are different from the real counterpart. Others are more adaptable and are able to look at a situation from the experimenter's viewpoint. There are pros and cons to both of these attitude styles.

The rigid, concrete-thinking mariner will give you unlimited advice on the realism of the simulation, the scenario and the quality of the actions of the bridge crew performing the exercise. Sometimes this can go too far, however, when the details being criticized are irrelevant to the experiment. For example, a mariner may say a houselight on the shore surrounding a harbor is missing or doesn't exist in the real situation. Depending on the importance of that light, the SME's opinion is either helpful or annoying. Another example, which actually happened, involved asking a number of experienced mariners to evaluate the potential training value of a particular bridge simulator. On the day the simulator was demonstrated, a number of minor problems occurred. The simulator froze a couple of times, and a bridge instrument did not work. The mariners were asked to ignore these problems and assume that for future training purposes everything would work properly. Some individuals were unable to do this and their evaluations were greatly affected by the problems they experienced in the demonstration.

Subject matter experts who are able to understand the simulation from the experimenter's viewpoint are helpful initially. However, they also adopt the biases of the experimenter and become advocates, rather than objective evaluators, of a simulator, scenario or training practice. In effect, they become "used up."

TECHNIQUES FOR CONTROLLING FMSs

In complex FMSs, it is very difficult to control the uniformity of a scenario when the participants' actions influence the course of events, which is most commonly the case. However, it is possible, within limits, to control the evolution of the scenario in various ways. Obviously, initial conditions can be the same. Other means of control are the assigned mission and contingency orders, and/or actions of other agents invoked to force the participant(s) to perform a desired action or choose a particular course. There is a trade-off

between maintaining the reality of the scenario and having events occur which channel the behavior of the participant.

Two tactics for controlling the scenario used by TH in maritime simulation are worth describing because of their implications for aviation simulation. The first is creating a subtle influence, essentially a conceptual barrier, that the participant will avoid. TH wanted the bridge crews to navigate through a cluster of other ships. To do this at some point the mariner had to choose to maneuver to the left. At the decision point the choice to go left or right was fairly arbitrary, although there were more reasons to go left rather than right. The mariner was not yet aware of the cluster of ships. Because the choice was not arbitrary some mariners would choose to go right and thereby unknowingly circumvent the eventual encounter with the cluster of ships later on.

In the original scenario there was open water to the right. To induce a left turn without being obvious, a sea wall or reef or the like was placed several miles away to the right. Because of the ship's location the scenario did not physically constrain the participant from maneuvering to the right, there was plenty of clearance, nor did the influence complicate the mariner's current situation. However, by moving in that direction, the mariner would eventually have to keep the barrier in mind during future maneuvering. So to avoid ever having to deal with the barrier, the mariners would consistently go to the left and eventually encounter the cluster of ships as desired by the experimenter.

Thus, the subtle influence of a potential problem or addition of another factor to contend with shaped the behavior of the mariner while still preserving the participant's freedom of choice. The important point is that the influence was of no immediate consequence nor would the influence be a significant factor even if the choice had been to go in the undesired direction.

The second means TH employed to control the navigation of the ship was the more conventional artifice of creating conflicting traffic. However, the particular technique for implementing it is worth noting. The conflicting traffic was a fast moving ship that would force a predictable, desired change of course. Normally in simulation, forcing events, i.e., actions by external agents or changes in the environment, must be preprogrammed to trigger under particular circumstances or at a specific time. In the case of the fast moving ship, it was programmed to maintain a constant relative bearing and heading to the maneuvering vessel and was located just out of radar range. That is, the ship was always lurking in the wings, waiting only for a cue from the experimenter to enter into the immediate gaming area. If it was not needed it was never called into play. However, if the experimenter saw that the mariner was not choosing the desired course, the fast moving ship could be brought on stage. Note that the experimenter did not have to enter course, range, and bearing information, which could be somewhat awkward during an experiment, but simply had to evoke a preplanned event that was dynamically tracking the vessel of interest.

COMPARABILITY OF STUDIES

MS mentioned that the scenarios she uses for the evaluation of navigational aids remain essentially the same for the important reason that it is necessary to be able to relate the results of past evaluations to current and future tests. That is, the need for compatibility forces the use of the same sce-

nario repetitively. The lesson here perhaps is that a choice of scenarios that may be used more than once should be done with great care or else at some future time a change in the fundamental character of the scenarios may imply losing comparability with a large body of accumulated data, or reluctantly accepting what is known to be a deficient set of scenarios for the sake of maintaining comparability.

APPENDIX E

FIELD INTERVIEW NUCLEAR POWER PLANT SIMULATION-BASED RESEARCH

David Woods
Senior Psychologist
Westinghouse R and D Center
Pittsburgh, PA

David Woods has conducted two studies on the design and utility of safety parameter display systems (SPDS) for nuclear power plant (NPP) operators. The purpose of his research, sponsored by the Electric Power Research Institute (EPRI), has been to evaluate SPDS design characteristics. His approach to this problem was first to determine how NPP operators would use an SPDS, e.g., as a diagnostic aid, as a planning aid, or as a control aid. That is, for what tasks and circumstances would an SPDS be an improvement over the conventional control room instrumentation?

For his research he used an NPP training simulator with the trainees, who were experienced operators, as subjects. The NPP simulator used is heavily scheduled six days a week. The experimental work was piggy-backed on the training exercises. This can be a very cost-effective means for MOR, if the experimental goals can be adapted to the on-going training situation. In the present case this was accomplished by using test events devised by Dave Woods instead of some of the training events. The trainees had no way of knowing the difference except for the fact that several additional observers appeared at odd times and for about half of the events the SPDS was turned on — a giveaway that an experimental trial was being conducted. The instructors were cooperative in allowing the experimental events to be used because they were very similar to the training events (thus fulfilling both the training and the research objectives) and because there was no grading of the trainees involved.

Nuclear power plant operators are required to undergo a week of training every year. The training consists of a mix of classroom work and simulator practicums. These are essentially problems involving a variety of emergency and out-of-tolerance conditions. The trainees are not required to pass a test or meet some performance standard. The requirement is simply to undergo the training. At the end of the simulator exercises, the instructor does debrief the trainees on their performance.

One of the principal problems in evaluating the utility of SPDS to NPP operators is gaining access to their performance. Their tasks are primarily monitoring and decision making. The overt behaviors associated with these tasks are frequently few and furtive. Information for commonly occurring or familiar simulations can be assimilated rapidly

and a decision about what to do, if anything, arrived at quickly.

Dave Woods used a clever method to reveal more about the information seeking and decision processes and also to improve the tractability of measuring performance. Several of the experimental events were designed to be slowly emerging situations. That is, even after the operator detected abnormal readings on some indicator, it took quite a while to get other information and formulate a tentative hypothesis of what was causing the problem and what to do about it. Even after some actions are taken, significant time is required to gain feedback information to confirm that the problem has been properly identified and the correct actions performed.

In addition to using slowly evolving events to reveal more about the behavior of interest, Dave Woods constructed events that superficially appeared to be a common failure to mask an unusual failure. This was most easily achieved by devising multiple-failure events. For example, a scenario event would include the failure of an instrument that, if operating, would signify that the event was an unusual and serious problem. The failed "leading indicator," however, led the operators to believe a minor, common problem was occurring, thus complicating and stretching out the process of identifying the problem. What would otherwise be a 2- or 3-min exercise would now run from 10-20 min or more. Choosing scenario events of this sort is a device for expanding the information seeking and decision time so that many more behavioral activities can be observed in detail. It is clearly a technique applicable to aircraft simulation research.

Subject matter experts (experienced NPP operators) were used to develop the experimental event scenarios, as well as to collect observational data during the tests and to interpret and evaluate the performance data. It usually took several iterations of refinement to get a satisfactory event scenario. For the two studies performed by Dave Woods, 12-event scenarios initially were developed. After preliminary testing by several groups of operators, he settled on seven events for the formal experimentation.

The primary data for the experiment were descriptions of the actions of the operators by other experienced operators. Dave Woods said it would be very difficult to obtain useful

automatically recorded data on the type of behavior involved in process control. In effect, the performance measurement system would have to be almost as smart as a human expert. In this case it was a lot cheaper to use a human expert. The automatically recorded data that were collected were in the form of a time record of the state of the NPP. The experimental events were also video taped for use during the performance assessment stage. Dave Woods refers to observational data collection as the ethnological approach to performance measurement, known as "watching the wolves mate."

During the first experiment there was no information on how the operators would react to the test events. After some experience of observing the responses of the operator, it was possible to develop a rating sheet for the SME observers to use in the second study. The rating form allowed for the fact that not all teams of operators would follow the same steps in dealing with the event. It included the full range of potential operator behavior. The observational and recorded data were collected and then assessed by other individuals. Dave Woods believes it is important to separate the collection of observational data from its assessment to minimize the loss of information from interpretive filtering at the collection phase.

EPRI originally suggested that observing the performance of operators in a high-fidelity simulation with the SPDS available, or not available, along with a few simple measures, would be sufficient to evaluate the SPDS. This, of course, is the common attitude of sponsors who assume that studies involving substantial behavioral components can be performed in a manner analogous to a shake-down test of equipment functions. A by-product of this expectation on the sponsor's part was that Dave Woods had considerable latitude in the approach taken in the first study. In effect, it was an exploratory investigation that served several purposes. Candidate experimental events could be tried out. The range of responses by the operators to the event scenarios was discovered as well as their apparent difficulty. Also, interesting leads could be picked out for more focused and more efficient investigation in the subsequent study. The first study also served to educate the researcher about how to conduct experiments in the NPP training simulator context. One of the most valuable lessons was on measuring performance.

The first experiment was fairly vague in purpose and simple in form. Dave Woods talked to several experienced operators to gain some ideas about how the SPDS might be useful. The events chosen for the experiment were typical of events used in other studies and thus could serve as a reference point to tie the results to other work. The operators were trained for only about 2 hr on the use of the SPDS prior to the study, so if anything the results would be biased for the conventional instrumentation conditions.

The conflict between the desires for face validity and experimental control in NPP simulator experiments is no different from any other simulator-based research. Dave Woods believes that any compromise should be in the direction of

gaining better control. He is willing to sacrifice some realism if the result is to drive or nudge the operators to the situation where the performance of interest occurs.

If the purpose of the experimentation is to derive information that is generalizable to other situations, there must be some concept that a condition in the simulated context was chosen to represent. A restricted or reduced simulation, perhaps with less physical fidelity, may have greater concept validity than an FMS. The reason for this, of course, is that a reduced simulation is designed to include only those characteristics that are essential to the concept. That is, the level of abstraction of the simulation approximates that of the concept.

Concept validity is very important in behavioral research; therefore less than FMS would be desirable for behavioral research in cases where the concept validity is improved by reducing the simulation scope. Although high fidelity FMS may have low concept validity, concepts must be finally tested in this context to prove they are robust and valid under near real-world circumstances. Most sponsors will expect research with some practical implications or importance to use FMS.

It is often very hard to convince sponsors or research result users that FMS is not necessarily the best way to do the research. Unfortunately, they are unlikely to get face-validity and convincing data out of a reduced research context. Dave Woods also stressed that it is important to be certain that the sponsor or user community has an answerable question. Frequently, they think an issue is clear-cut (e.g., is an SPDS useful); but, from an experimental point of view, it is only a vague notion. It is usually necessary to refine the question to some tractable form.

Realism in NPP simulators in terms of user acceptance is not as serious a concern as in many other forms of simulation. The training simulator used by Dave Woods is a high-fidelity physical and functional representation of an NPP control room. Any issue of fidelity or realism would occur only if the simulator did not respond as a real NPP does or if the scenario events did not have a plausible basis in reality. As Dave Woods pointed out, most experienced operators have never been involved in a real abnormal event and, because of the large number of possible failures, the operators have no idea before or during the course of an event whether or not it is realistic. After the exercise is over and the precipitating cause of the event is explained, then and only then are the operators able to comment on the realism of the event and its chain of effects. This is obviously of great benefit to the experimenter. The researcher has wide latitude for composing event scenarios to serve the experimental purpose. Incidentally, the operators become very involved in the scenarios and at the end are very interested to find out what the "cause" of a simulated event was. Dave Woods pointed out an outstanding issue about use of reduced fidelity simulation for NPP research; no one knows what the

attitude of experienced NPP operators would be to reduce fidelity simulation because it has not been tried.

Similarly, the inability to assess what is real could be the case for pilots of future commercial aircraft. New generation aircraft and NPP are both examples of supervisory control systems in which ongoing processes are automatically controlled. The operator only monitors the process and usually only intervenes under abnormal circumstances. Like the NPP operators, there is no way that a pilot can know all of the possible ways the automated systems in future aircraft might fail, or know all of the possible indications and consequences of its failures.

There was one realism problem in the experimental event scenarios used on the training simulator. All desired effects such as failures have to be scheduled by time, and not by contingency on some specified set of conditions. One event involved the failure of a valve. In the real world it is plausible for a valve to fail under pressure or when it is being opened or closed; it is not realistic for it to fail while in some benign, quiescent state. Yet, in the simulation, because the valve failure had to be scheduled by time, it would appear to the operators that the failure was capricious if there were no circumstances at the time of failure which would plausibly be

associated with the failure. This obviously detracts from the realism of the simulation.

Valve failure on time alone also had a detrimental consequence for the experimental control. Because the valve failure occurred on a time basis and was unrelated to whatever the plant or the operators were doing, it would happen at a different point in the chain of actions for each group of operators. Thus the conditions were not exactly comparable from one group to the next. It is a relatively small point, but any research simulator should have the capability to permit scheduling of experimental events contingent on circumstances and subject actions rather than time alone.

Dave Woods believes that there are some marked differences in the behavior of NPP operators in simulators and real plants. In simulators, operators are much more action oriented; that is, they are more willing to take some positive step in an abnormal situation in the simulator than in the real world. He attributes this difference to the fact that in the real world operators are very concerned about repercussions from any actions they may take. The same sort of stress does not arise in training simulator exercises. However, he did say that there appear to be no obvious differences between actions taken in the two settings.

APPENDIX F

FIELD INTERVIEW THE BOEING COMPANY

Thomas C. Way
Avionics, Crew Systems Department
The Boeing Co.
Seattle, Washington

Dr. Richard E. Edwards
Managing Consultant, Human Performance Analysis
The Consulting Division, Boeing Computer Services Co.
Renton, Washington

Thomas C. Way and Richard E. Edwards have been involved in the development of simulator scenarios for evaluation of cockpit equipment. Tom Way did the simulator evaluations of the Boeing-proposed version of the C-14 intended to replace the C-130. The principal issue in the evaluation was workload since the C-14 concept featured a two-person cockpit crew instead of four persons as in the C-130 (pilot, copilot, flight engineer and navigator). Tom noted that the C-14 concept development was the first major program that included human factors members as part of the design team rather than as part of logistics support.

Richard Edwards until recently worked in the crew systems divisions and has conducted a number of simulator-based developmental and FAA certification evaluations for military and commercial aircraft and systems, including the B-757/767.

Boeing has three types of aircraft simulator facilities, 1) flight crew training; 2) developmental, which range in quality from rudimentary to near flight-training fidelity; and 3) engineering, which are readily reconfigurable but have reliability problems.

Analysis, simulator-based performance tests, and questionnaires given to experienced pilots are the three main ways of evaluating new equipment or systems. These methods are applied extensively for workload determinations. It is one of the principal concerns in commercial and military aircraft development.

The C-14 workload evaluation started with an analysis of imposed workload (task load) for various situations. The focus was on the crew's ability to perform navigational tasks as well as piloting tasks. Eliminating the flight engineer was never considered to be a problem. The flight simulation scenarios were 35-45 min segments, and the locale represented was in the vicinity of the Rhein-Main AFB, Germany. Ten C-130 crews were used for the evaluation; each crew was available for only 2 days. The first day was devoted to training, primarily on the use of a control and display unit (CDU) for navigation, since its use would be a critical factor in the

simulation evaluations. The second day was devoted to testing.

The simulator was a fixed-base cockpit with a monochrome projection visual display on a 16-ft screen, 12 ft from the cockpit. The visual system was used only for takeoff and landing. On takeoff the aircraft would enter weather, and the remainder of the flight would be on instruments. A fixed-base simulator was considered to be appropriate because the flight-control stabilization system made aircraft control relatively easy. In military cargo aircraft such as the C-130, a normally difficult flight task is low-level cargo extraction because there are sudden significant changes in the center of gravity as the cargo is discharged. The proposed stabilization system would automatically compensate for these types of changes, so the extraction task was not included in the scenario. Aircraft handling was not considered to be an issue. Also, a Boeing test pilot working on the C-14 project said a fixed-base cockpit would do just fine since the motion base gives the wrong cues anyway.

A very good model of the aircraft dynamics was included in the simulation. Most of the cockpit instruments, which were of conventional electromechanical design (other than the navigational CDU), were functional.

The scenarios centered on creating navigation problems. The crews were required to accept in-flight diversions and do such tasks as estimate fuel states and insert way-points using the navigation CDU. The object of the study was not to compare alternatives but to confirm that the crew size and cockpit-design concepts were practical.

The operational question was whether the crews could perform the tasks without feeling excessively burdened. A modified Cooper-Harper rating was used to assess workload. The test scenario would be frozen at a natural task breakpoint and the crew asked to rate the segment just completed. A natural breakpoint was considered to occur when the diversion or similar problem requirements had been completed and the aircraft was beginning a cruise segment. It was considered more desirable to disrupt the continuity of the

scenario to get immediate ratings than to rely on memory after the scenario was completed.

Tom commented that the evaluation tests were planned and executed very quickly after the preliminary design concepts were formulated. The object was to have results in sufficient time to influence the final design proposal. He said, also, that a considerable amount of part-task simulation work was performed in the development of the navigation CDU. Therefore the mission simulations were not to evaluate the navigation instrument per se, but to evaluate its effect on the overall mission workload.

At this point, Tom Way was called out of the office to be told he had to travel to Wichita, Kansas the next morning. Since he had several matters to attend to, the interview was cut short before he had an opportunity to make some general observations on the construction of simulation scenarios. Richard Edwards, having heard Tom describe the C-14 evaluation, did not describe in detail any of the particular simulation efforts he had directed but offered several observations on the use of MOS at Boeing for human factors developments and evaluations. His comments are as follows.

APPROPRIATE LEVELS OF SIMULATION FOR RESEARCH AND EVALUATION

When a new full-mission simulator is installed, there is initially low demand for its use; consequently, everything is done in the simulator even though a part-task or laboratory study may have been more appropriate. Later, when the simulator is in heavy demand, the issue of what studies should be done in the simulator becomes one of more concern.

There are four principal reasons for doing FMS:

1. To resolve a collection of related problems. If there is a series of part-task evaluations called for that are related, for example, evaluations of several different instruments and controls for the same aircraft, it can be more economical to gang them together in a comprehensive study.
2. When the focus of interest is on long duration or infrequent effects and events. Behavior under fatigue and responses to rare emergencies as a function of time-on-duty are obvious examples.
3. Subtle interactions may influence the behavior of interest. Results of crew coordination studies are likely to be adversely affected if the simulation is not physically comprehensive and realistic, or if the scenario is too short.
4. To evaluate performance of people and/or equipment that occurs during a series of transitions from one flight phase or mode of operation to another. For example, MOS would be appropriate to evaluate performance when display formats change during the phases of a descent to landing.

However, sometimes a simulation requirement can become so complex that it is better for overall efficiency to subdivide the effort. For example, in the DAIS project, there were five separate laboratories working on separate parts of the instrumentation problem. Each had a model or computer simulation for their particular function which was fed to a single simulator cockpit. Problems with the individual components resulted in an overall Mean Time Between Failure (MTBF) of less than 1 hr. They eventually gave up the pursuit of an early, integrated evaluation of the new displays and functions.

Part-task simulation is appropriate for research or evaluations when there is no reason to suspect behavior will be influenced by secondary contextual circumstances. Generally, studies intended to determine what the best performance can be in specific conditions, or to discover whether or not inherent functional problems occur, are the proper domain for part-task simulation. In addition, if research is focused on a particular problem, do not add extraneous things.

The appropriate level of simulation realism and comprehensiveness for a particular problem is usually obvious. There are few instances of gray areas where the characteristics necessary for the investigation are uncertain.

SCENARIO DEVELOPMENT

Stock scenarios tailored to the particular study are often used at Boeing. The experiment or evaluation is first designed at the conceptual level and then SMEs are used to help develop specific scenarios that embody the concept and also appear realistic to the experienced participants. Subject matter experts are very valuable sources of detail for scenarios. For example, in one study an approach was to be flown to an airport using a particular runway for day and night landings. The expert pilot consulted about the scenario pointed out that the selected runway was not used at night because the winds always change from day to night.

The procedure for scenario development should start with a time-line description of the tasks to be performed by phase of mission, and the hardware, software, or system events that will be the conditions of interest or of the experimental variables. The scenario should avoid any cascading effects. That is, a test event should not occur too soon after a previous one to avoid interaction effects unless they are desired. Pilots or other crew members will worry about the last event for periods of time, particularly if it resulted in an unresolved problem. If this is likely, there should be a period of stable operation or settling down before the next event is encountered. Subject matter experts, experienced pilots, can be helpful in determining the appropriate sequence and timing of the critical events.

CONTROL OF SCENARIO

Scenarios can be designed so that the subject's behavior is either tightly controlled or allowed considerable latitude depending on the purpose of the study. In most evaluations, adherence to a desired sequence of performance is sought because the intent of the evaluation is well defined. More open-ended scenarios are useful for exploring for problems or seeking the range of possible alternative actions. The latter is less likely to occur in systems development in industry. It is more likely to be a goal in government or other research facilities.

In industry applications of simulator-based evaluations, it is generally desirable to gain as much control of the study as possible because of the few subjects that are likely to be available. Most studies involve 10 or fewer test subjects or crews. Therefore, every effort is made to preclude deviations from the desired course of events and reduce the performance variability among the subjects or crews so that differences in the conditions of interest can be detected. Compliance to the desired scenario profile is attained by briefing the pilots or other test subjects on what is expected, and subject variability is reduced through pretraining in some circumstances.

Pretraining should not involve the experimental task directly but should develop the basic skills required and familiarity with the equipment operation. Training should proceed to some predetermined criteria to minimize performance differences in the experimental task. Pretraining is not always possible when complex behaviors are involved but it is feasible for procedural and skill-type tasks. For example, a study was conducted to evaluate several alternatives for locating certain buttons and switches on the hand controls for an aerial refueling boom. The same subjects were used to test the various configurations. The simulator test involved making contact with a receiver aircraft that was preprogrammed in flight profile. To minimize inherent skill differences and dissipate proactive interference effects among the configurations, the subjects would practice before each trial with another set of controls. The practice task was touching the boom tip to designated squares of a checkerboard. The practice would continue until the subject could perform this task to specified criteria of time and accuracy. Only then would they enter the simulator to perform the experimental task.

DATA COLLECTION

Automate data collection and analysis as extensively as possible. It results in a rapid output of information, and avoids errors common in manual collection and transcription of data. It is common practice to use multiple redundant

data recording for particularly critical data; that is, two digital data recorders are routinely used to log data. In the Boeing simulators, it is possible to get a full time/event history. Initial switch settings and system states are logged and changes are identified and time marked.

Also, data are collected in several forms. Video recording is used whenever possible. In the B-757/767 evaluations, audio and video recordings were time marked for comparison with the event record to determine differences from the desired profile. The video tapes are particularly helpful during the debriefing of the pilots. A situation can be replayed to refresh their memory and prepare them to answer questions or make evaluation ratings. In some cases the several forms of data collection are different transformations of the data. For example, a ground track plot and a time to perform record was made in a study comparing manual vs. automatic VOR tuning. The time plots showed no differences, but there were dramatic differences in the ground tracks (automatic tuning was better).

Observational data from experts are used only rarely because of concerns about differences in interpretation. However, postflight questionnaires, rating scales and debriefing interviews are used regularly. One problem with preference ratings is that a less-preferred condition will sometimes result in better objective performance.

Preliminary testing of data acquisition and analysis routines is essential. Performance measurement requires a good deal of software development and intentions are frequently misinterpreted. Checking of these routines must be intensive; many surprises are usually found. The performance measurement and analysis testing involves testing it yourself with known data, running in-house subjects, and running preliminary subjects from the population of interest.

ATTITUDES OF PARTICIPANTS

Pilots, both military and civilian, are very tolerant of shortcomings in simulators as far as the physical characteristics, visual scenes and aircraft dynamics are concerned because most pilots today have had long experience with simulators and understand their shortcomings. However, they are very intolerant of unrealistic procedures, events, and conditions. Some pilots, probably because they have engineering backgrounds, are troubled by equipment evaluations for technologies that do not exist. For example, a study of display formatting and other characteristics for information from a sensor or processor that does not exist gives them a great deal of trouble, even though the issue is the information presentation, not the source.

Pilots also are generally very accepting of the purpose of a simulation exercise, and what is expected from them. Interestingly, the most difficulties occur with Boeing pilots who participate in many simulation studies. It may be that they

view participating in simulations as secondary to their primary duties of flight testing. Also, the simulation studies are often scheduled at night, and their participation amounts to unpaid overtime.

The pilots are always asked to comment on the simulation and scenario to find out if they feel the simulation was appropriate to the purpose. Sometimes the structuring of the scenario can go too far. In one case, a script for the pilot to read as a passenger briefing was provided. Many pilots felt this was a bit much, particularly since most said it was not exactly what they would say in an announcement.

EDUCATION AND RESPONSIBILITIES OF THE SIMULATOR RESEARCHER

In complex simulator-based experimentation and research there is both an experimenter learning curve and an experimenter teaching curve. The learning involves gaining an appreciation of how difficult it is to plan and execute a major simulator study and how meticulous you must be. Most of the effort occurs before and after the actual data collection, i.e., while running subjects. About 60% of the effort is in preparation, 30% in data analysis and interpretation. Only about 10% of the effort is the execution of the simulation runs. There is no apparent procedure for determining how difficult a particular study will be.

The teaching curve is manifested by experienced researchers imparting to less experienced or novice researchers the many problems to be aware of, and the relative importance and effort required for, various aspects of preparing for and executing simulator experiments and evaluations. Every large-scale simulation facility should have a cadre of experienced researchers to assist in the conduct of studies by colleagues from other divisions who may be responsible for conducting a simulator study. Some of the points, already mentioned earlier, that are not fully appreciated by naive researchers is the need for detailed planning, a great deal of continued consultation with experienced pilots, extensive shakedown testing, verification of the data acquisition and performance measurement system, and the absolute necessity for comprehensive preliminary testing.

Probably one of the most difficult problems of the first-time experimenter is becoming familiar with the physical design and operation of the simulator and the role of the facility support personnel. The last is especially important because the experimenter may have expectations that these personnel understand the intent of the study, what it implies in terms of preparation of the simulator, and what they must do to support the study. This is not likely to be the case. Requirements, down to what the initial switch settings in the cockpit must be, and who is to set them, must be stated explicitly. No one but the experimenter is going to worry

about details, and he or she must be certain that none have been missed.

An occurrence at Boeing is a good illustration of what can happen if the experimenter is not thoroughly familiar with the simulator facility and personnel. An evaluation of display concepts using experienced pilots was scheduled for several nights running. The results were needed urgently. The preliminary preparations were made and the display hardware and software installed in the simulator. The evaluation was to be based on observational data and the opinions of the pilots. Shortly after the test began, the computer locked up. This occurred repeatedly, but never at the same time or during the same events. Several hours were spent, on several successive nights, looking for the programming error that seemed to be the most likely cause of the lock-up. Of course, much valuable time was wasted and there was a loss of the scarce pilot resources.

It was eventually discovered that the cause of the failures was a tape-drive write-error. The simulator had a built-in data logging system, but it was not being used in this particular experiment. The support personnel knew that a tape was supposed to be mounted onto the drive to log the data, but this particular experimenter did not provide one, contrary to normal practice and the expectations of the support personnel. Therefore one of the facility people put a discarded tape on the drive so the simulator would run. The researcher was unaware of all this, and the well-intentioned support person had no idea the tape was the source of the problem. It turned out, of course, that the tape was discarded because it was bad, and when it was used on the drive, a write-error would occur at some random point.

The researcher is equivalent to a general building contractor. He or she must know the simulator operation and capabilities, and the facility management and support personnel. The researcher must take all responsibilities for the planning, issuing of instructions, coordinating of support requirements, and checking and verifying of software, hardware, and procedures prior to the study, as well as conducting the data collection and performing the analyses.

ATTITUDES OF CUSTOMERS AND USERS OF INFORMATION FROM SIMULATOR STUDIES

The customer's expectations, whether the customer is an in-house user or an external organization, have a marked influence on the form of the simulator-based tests and evaluations. They tend to overly stress face validity as a requirement, and often expect more to result from the tests than is possible. Engineers especially have expectations that behavioral information should be readily available, or can be quickly acquired to answer their questions. Engineers frequently confuse experimentation with demonstration, and they do not fully appreciate the importance of good

experimental control as a factor in the validity and reliability of the data. Accommodating the biases of the customer and user to some degree is necessary. It is also possible sometimes to explain why certain methods are important to the goal of the test and why certain features of the simulation are not very important.

PLANNING FOR DEMONSTRATIONS

Outsiders, particularly, managers, marketing representatives, senior personnel from the client organization, and other VIPs do not accept the importance of not interfering

with the conduct of an experiment. They expect to see demonstrations and think little of entering a cockpit during a run. Unfortunately there is little that can be done to discourage this practice. The wisest policy is to recognize that demonstration is an important part of any development program and serves legitimate needs even if it is inconvenient and sometimes seriously detrimental to the experimental plan. One of the best means to avoid interference to the extent possible is to plan for demonstrations before and after the data collection period, allow for it in time and budget plans, and be sure key people are made aware of the availability for demonstration well in advance. This will not stop interference but it will help.

APPENDIX G

FIELD INTERVIEW SIMULATION STUDIES OF VISUAL ILLUSIONS DURING NIGHT APPROACHES

Dr. Conrad L. Kraft
Bellevue, Washington

Dr. Kraft retired from the Boeing Company in 1983. During his long tenure with the company he performed numerous studies of pilot vision and performance, safety, and simulator visual characteristics. He received numerous awards for his simulator-based research relating illusions of altitude during night visual approaches to the visual and geographical characteristics of the airport and surrounding area. His basic finding was that an illusion of excessively high altitude was manifested when an airport is situated at the edge of a city and there is an upward tilt of the city from 1° - 3° from the airport to the horizon. The pilot of an aircraft making a straight-in, night, visual descent from a high altitude over water or otherwise dark terrain toward the airport with the tilted city in the background will misperceive the aircraft altitude to be much higher than it is. The consequence of this illusion is the aircraft will contact the ground 5-8 miles before the threshold of the runway.

The research was prompted by a series of crashes of B-727s within a few years of the introduction of the aircraft into commercial service. Dr. Kraft was asked to determine if there were any characteristics of the aircraft that could be contributing to the accidents. He discovered through looking at the descriptions of accidents involving the B-727 and other commercial aircraft, that many of the accidents occurred under conditions of approaches at night over dark areas to airports near cities with an upward tilt.

Dr. Kraft was struck by the common circumstances surrounding this large proportion of accidents and the strong suggestion that vision was involved. He began a series of investigations on the visual perception of altitude. He began by photographing maps of some of the cities where the accidents occurred. He placed glue on the maps outlining the runway and airport and roads of the city plus random spots in built-up areas. He then sprinkled fluorescent chalk dust on the maps. The maps were photographed under black light from various distances and angles, corresponding to a long descent path to the airport. The set of photographs appeared as night scenes of an approach to an airport. Three groups of pilots, current airline pilots, noncurrent airline pilots, and noncurrent small aircraft (low altitude) pilots were asked to sort the photographs into altitude bins. The result was the first group was reasonably consistent in their judgments but the last two groups showed high variability. The lesson from this study was that only current airline pilots would be appropriate subjects for future studies.

The next study also involved judgment of altitude based on viewing 16-mm moving films of a model city. The camera moved down tracks set to represent two high and two low approaches. The judgment data from the airline pilots was inconsistent. The film resolution was not high enough to give an adequate representation of point light sources as seen in the real world.

For the third study, a large model city mounted on a moving table was constructed. The budget limit for the equipment was \$12,000. The model was made by puncturing pin holes through a large print of an aerial photograph of a city and airport. An important detail was to puncture the holes with a soft wood backing under the photograph so that each light point was dimpled. This was necessary to preclude the whole scene from suddenly going black if the simulated aircraft went below the plane of the table. The slightly raised holes were in effect small spherical light sources.

Prior to construction of the city model, an analysis was made to determine if the model moving toward the pilot in a mocked-up cockpit would provide any monocular cues to its true distance. The vision literature indicated that accommodation, the only available cue because the pilots viewed the scene with one eye occluded, would not be effective until the table came within 28 in. of the observer. In effect, the city could simulate an approach from 20 miles out, to within 4.5 miles of the runway over altitudes in excess of 20,000 ft to below ground level. The city could be inclined up to 3° . The runway remained horizontal. The cockpit first used was a left-over, single-seat fighter mock-up. The aerodynamic model was a nonspecific representation of an aircraft weighing approximately 100,000 lb. The only instruments available to the pilots were airspeed and vertical speed indicators. The aircraft did not have horizontal, yaw or roll movement.

The Boeing pilots objected to the absence of an altimeter, believing they could not make a proper descent and approach without it. They were told the object was not to evaluate their performance but to determine if the city simulation was adequate. This allayed the pilots' concerns. As it turned out, the pilots could fly the descent very well with only the two instruments and the visual scene. After the experiment, the pilots were all very surprised to find that the city moved toward them rather than vice versa. It was a good confirmation of the absence of extraneous visual cues.

Additional funds provided to the project because of a need to move the laboratory allowed the simulator to be upgraded. A cockpit resembling a transport aircraft was constructed, and the aerodynamic equations were improved to represent an aircraft with a gross weight of approximately 150,000 lb — somewhere in the weight range between a B-737 and a B-727. The pilots had been extremely critical of the handling characteristics of the earlier simulator. With the improved aerodynamics, there was a marked reduction in the variance of the glidepath for each pilot.

Pilot acceptance of the simulator helped the study because word was passed around that it was a good simulator. This considerably eased the problem of recruiting pilots to participate in the study. Another factor which prompted cooperation by the pilots was that almost all of them had had a close call descending below the glideslope at night. The enthusiasm of the pilots about the importance of the study also influenced management to continue support of the study.

Because of the limited task requirements, control of airspeed and descent rate, Dr. Kraft was concerned that the imposed workload was so low compared to actual flight that the pilots may be concentrating far more on altitude judgment than they would be in actual operations. To increase workload other air traffic was added to the simulation. That is, the pilots saw aircraft beacons flying over the city (the beacons were actually small lights mounted on a few rotating disks ganged together). Because of the radii of the disks and their slow movement they appeared to the pilots to move in a straight line. The pilots were asked from time to time to report on the azimuth, heading and relative altitude

of the other aircraft (via a request from an air traffic controller). This side task both increased workload and drew their vision away from the approach task as a normal scanning of the airspace would do. The pilots were also asked for their estimated altitude at precise points in the approach. This was accomplished by computer control of a tape recorder. A nice touch was that the voice on the tape was that of the person playing the role of the air traffic controller who asked for the information on the other aircraft. Thus, it was not apparent to the pilots that a tape recorder was making the requests.

Reflecting on the series of experiments, Dr. Kraft made several noteworthy comments. Maximum effort was placed on visual-scene fidelity because of the suspected perceptual nature of the problem. Fidelity of other features were of minimal concern because they were judged to be of little consequence to the characteristic of the performance of interest — going below the glideslope. The very limited funding for the study forced concentration on the most critical feature of the simulation. Had a full-mission simulator been made available for the study, Dr. Kraft said they probably would have not discovered the basic problem. Conventional model-board systems simply could not portray the characteristics of point light sources and thus cues responsible for the illusion would be absent. This study evolved through a number of stages deliberately. Dr. Kraft said that pilot experiments are really very necessary to help you think about a problem. He ventured that anyone who attempts to perform an FMS experiment without having conducted several preliminary studies of more modest scope would be very lucky to have a successful outcome the first time.

APPENDIX H

FIELD INTERVIEW SIMULATION AT THE UNITED AIRLINES FLIGHT TRAINING CENTER

Dave Shroyer⁴
United Airlines Flight Training Center
Stapleton International Airport
Denver, Colorado

Dave Shroyer has been involved in United's training programs since the 1950s, when they had fixed-base instrument trainers in Chicago. His primary thrust since 1952 has been to advocate recurrent proficiency training. United started line-oriented flight training (LOFT) in 1976, and more recently has developed command, leadership, and resources (CLR) training. Dave has a group of about 35 people including Flight Standards instructors and analysts to develop and execute training programs for all six aircraft that United operates.

LOFT TRAINING

Flight simulator currency training (recurrent proficiency checks — PCs) can be divided into "batting practice" and LOFT. Batting practice represents the traditional approach — successive approaches, departures, failure modes and the like, not in a trip context. Dave emphasized that they do not slew or restart the simulator at initial conditions (the beginning of an approach), but fly it around the pattern. Flying large jet aircraft requires staying ahead of the aircraft and anticipating future events. He felt that pilots lose the context and pacing of normal operations if they are slewed around to the start of an approach. [Note this is a different position than was taken in the initial B-767 Computer-Based Training system by others at United, but the automated B-767 training program did not materialize, primarily because of a last-minute change from three-man to two-man operations and the impact on the simulator redesign, delivery schedule, and costs.]

LOFT is well known as training in the context of a line trip from point A to point B. LOFT objectives are to provide training that combines the aircraft, the route, and crew interactions within cockpit and between cockpit and all external systems such as dispatch, ground crew, and ATC. LOFT provides all normal trip activities, including trip paperwork, for the whole crew. It permits crew interactions and exposure to past and recent line operational problems. Given the proper selection of valid United line trips, many of the problems which really occur can be built into the scenario. Dave claims

that if they are given more time, they could work all of the batting practice drills into a LOFT in certain areas, such as Southern California, Chicago-Cleveland-Pittsburgh, and Boston-New York-Washington. Current proficiency checks have two segments of LOFT in addition to the customary Batting Practice.

CLR TRAINING

There is heavy emphasis on CLR training because 80% of the accidents have nothing to do with the aircraft, but with "human factors," which also used to be called "pilot error." Dave stated it was common knowledge that prior to CLR, First Officers would become Captains on the basis of time alone (seniority) and flying proficiency (being able to pass the checkride). They were thrust into command and decision-making jobs without the benefit of command and leadership training. He also observed that the need for such training was not limited to Captains, because the whole crew operates the aircraft, and must function efficiently and safely.

CLR training is composed of academics (learning the theory from text and workbooks), seminar discussions, and specially designed and debriefed LOFT exercises. CLR emphasizes the concepts of a) Inquiry, b) Advocacy, c) Conflict Resolution, d) Decision-Making, and e) Critique as elements in the identification and resolution of operational problems in the cockpit. The approach uses the Blake and Mouton (1978) Grid to identify individual pilot styles. Blake and Mouton classify individual behavior in terms of either a basic task or people orientation. They quantify these two attributes in an x-y matrix scaled from 1 to 9 on each axis. Pilot styles can be identified and scaled along these dimensions. Dave commented that the military (or "captain is king") style does not promote the best use of cockpit resources; actually, Dave was more emphatic when he said, "It doesn't work."

The key elements of CLR training are as follows: a) there is no single solution to the problems given, b) there is no interference by instructors, and c) there is no performance assessment by instructors. In a typical LOFT exercise for

⁴Now retired from United Airlines.

CLR training, a major problem, which has a cascading effect, will be introduced early in the flight; problems such as major hydraulic system failures are used. Minor "mosquito bite" type problems such as light bulb failures are avoided.

There may be more than one problem, but they typically do not make the scenario too complicated; the normal constraints of flight and the ways crews interact to solve problems create a fertile environment for CLR training. The flight is permitted to develop naturally. No one interferes, stops the flight or tells the crews what to do. All scenarios are possible to execute safely, but if the crew makes too many mistakes, or a critical mistake at a wrong time, a crash could result. A crash is permitted, but Dave said they have never had one.

United always provides at least two viable courses of action, so that the crew will be forced to make decisions, and will be able to use innovative problem solving techniques instead of being led down the path of least resistance by the constraints of the scenario. This echoes concerns that some scientists have had to permit truly emergent behavior to unfold in team training situations (Crowe et al., 1981).

The crew is video taped, and all conversation and communications are recorded. At the conclusion of the flight, the video tape is taken to a closed room for crew review and debriefing. The instructor serves as a facilitator only; he does not evaluate or offer comments. He leads the discussion and focuses it to particular parts of the flight that the crew should review and critique for themselves. At the conclusion of the debriefing, the video tape is erased.

COMMENTS

Problem Selection

For LOFT and CLR, United looks for problems which are realistic, solvable, and have multiple implications for the remainder of the flight. As said before, they avoid "mosquito bite" problems, and look for those which will tax the teamwork, system knowledge and decision making capability of the crew. They choose problems which will cause the crew to think ahead, perhaps to the approach and landing, and to plan what they would do if there are further problems, such as other failures, a change in weather, or a change in the landing runway. Problems can include but are not limited to the following types:

- Electrical, hydraulic, and mechanical system failures.
- Flight paperwork errors, dispatch procedures, weight and balance.
- Crew or passenger problems (including bomb and hijack threats).
- Problem ATC, noise abatement, or obstacle clearance procedures.
- Problem airports, landing runways, and traffic delays.

- Weather and effects on takeoff, enroute, fuel, and landing requirements.

United derives data from accident reports, from irregularity reports within the company, and there is a "network" among the airlines and equipment manufacturers to share operating problems and solutions. United trains about 6,000 pilots a year. Their instructors receive constant feedback on what happens on the line and, of course, they gain information on potential operating problems in the simulator training sessions. In addition, there are line-check pilot reports and quarterly flight standardization meetings. In short, there is a constant flow of information on equipment, maintenance, ATC, airport, dispatch, route, and crew difficulties. The severity and implications of problems (most of which actually have occurred in line operations) along with a judgment of the ability of training to mitigate the problems, drives the selection of problems for LOFT or CLR training.

Using these data, Dave Shroyer, who has more than 30 yr of experience with what does and does not work in training, collaborates with an aircraft fleet representative and assistant. United has six aircraft types. A fleet representative and assistant represent the technical expertise on each aircraft. Dave frames the problem generally, the fleet representative and assistant write the training objectives, and Dave reviews their work as a quality control check. Thus, three people are directly involved in problem selection and scenario definition at the level of the training objectives. It must be remembered, however, that problem selection is based on a cafeteria of data and information which has been derived from all the sources which were described above.

Scenario Construction

About 4 wk is allocated to the construction of a new scenario and three or four people are involved. This time allocation assumes that the scenario analyst knows the aircraft and systems very well, the normal and emergency procedures for the aircraft, the flight operations procedures used by United from dispatch to arrival at the gate, and the details of the specific departure airport, the route, and all possible terminal areas and airports.

The estimated level of effort assumes that all the required data for frequencies, facility locations, terrain and airport models reside in the simulator data base. It also assumes a knowledge of candidate problems to give the crew, and the training objectives that are addressed by those problems. Dave is hesitant to guess what level of effort would be required to build a scenario from scratch without all this institutional memory. He believes at least 6-8 person months would be required, and possibly more, depending on the experience of the scenario development team.

Once the departure airport, route, and nominal destination are determined, the actual line route is observed by one of the scenario designers (in the jump seat), and all radio and

intercom communications are recorded from pre-push back to arrival at the gate. Air traffic control personnel at departure, enroute, and arrival locations are consulted. This is done to ensure that the actual language, procedures, and flow of events of each route are as realistic as possible, and up to date. The scenario is then built and reviewed by Dave to ensure consistency, sufficient pacing of events, meeting of training objectives, and realism. If the training is to be certified, the FAA has to approve it.

Scenario Testing

About a week is devoted to preliminary scenario testing to be sure that it works in the simulator. Flight Standards instructors at the training center review the scenario, both from an execution viewpoint in the simulator and to test the utility of the scenario for its training purpose. After preliminary testing, the scenario is modified as necessary during initial instructor training. Scenario testing continues and it may be revised further after the first crews fly it. Flight crews often point out improvements in young scenarios. Scenarios are not considered to be debugged well or relatively stable until they have been used for about 6 mo. Even after that time, they have to be changed if there are any changes in the routes, frequencies, facilities, or procedures.

Instructor Training

Instructor training to administer a scenario requires about as much time as building the scenario initially. The nominal time is about 4 wk. United instructors maintain currency in all positions on the aircraft they are training, and maintain currency on more than one aircraft. They have a reservoir of institutional memory of prior scenarios and the general principles of LOFT and CLR training. Instructor training time could easily quadruple the current time allocated if this level of currency and institutional memory were not available.

Cabin-Crew/Flight-Crew Interactions

Communications with the cabin crew can be simulated easily, as can communications with the ground crew. If necessary, United will simulate sending a crew member "out of the cockpit" if the problem dictates. The rationale is that, for the training problem (and in reality), that person is simply not in the cockpit during that time. He or she will return and make a report.

For the crew remaining in the cockpit, the illusion of flight is preserved. For the person who "leaves" the cockpit, the illusion may be interrupted, but he or she is doing what they would normally do — leave the cockpit to do something and report back. The overt behavior is consistent with

reality. United doesn't seem to be concerned with whether or not the illusion is maintained for the person who "leaves" the cockpit. The problem itself is real, and the solution provides crews with the experience of dealing with it, which meets United's training objectives.

Stress and Peer Pressure

As for stress created by peer pressure, it was observed that most professional pilots are quite sensitive about their performance. It was doubtful that being in a simulator with other crew members and an instructor would change this source of stress from what normally exists in the real world. Since United is dealing with proficiency checks which have a direct bearing on continued employment, it is possible that there is more stress in their simulators than on a normal line flight.

Organizational Pressure

As one approaches LOFT-type scenarios and is investigating the crew interactions and decision making that might occur in the real world, there are organizational pressures that might influence behavior in the real world, but may or may not influence behavior in the simulator.

For example, if one declares an emergency, a report has to be written, and the problem becomes known to many people in the company hierarchy. Some pilots may not want the hassle of the report, being "second guessed" by someone who was not there to see the whole situation unfold, or having their name associated with an operational problem, however mundane. As another example, there are difficulties for the company and the passengers if an aircraft does not land at the intended airport, or one which can handle the flight, maintenance, or passenger requirements.

Pilots are trained throughout their careers to maintain a reasonable margin of safety; but it is seldom that an aircraft is dispatched in perfect working order, that all facilities along the route are operational, or that the weather is certain. Pilots have to make judgments. There is pressure to make each flight as economical as possible, and there are relatively few absolute criteria. Pilot behaviors, and especially their decisions, will be conditioned by organizational pressure in the real world. The extent to which this behavior is exhibited in a simulator is unknown, but probably varies from pilot to pilot.

Dave Shroyer commented that United has not begun to address pilot judgment and decision making directly. Instead, they develop scenarios which will challenge the flight crew teamwork and decision process, let the flight unfold without interference, and guide the flight crew in their own critique. No judgments of the goodness or badness of the performance are made by the flight instructors during CLR training.

Nonpilot Observers

We asked Dave if a researcher could see what a skilled and experienced pilot would see while observing a simulator flight. He commented that one designs a simulator flight with some purpose in mind, and it doesn't take a skilled and experienced pilot to determine whether or not the desired behaviors resulted. Observers have to be trained what to look for, and they may need training to increase their observational skill for a given environment and situation. Undoubtedly, flight training and experience would help the observer understand the environment and what to look for, but, given some training, it was not essential. Dave commented that one of his scenario analysts is a psychology major, but he designs very good scenarios and understands the cockpit environment very well.

Fidelity

We had to ask the obvious question: Can some or much of this training be done without expensive, high fidelity simulators with visual and motion systems? Dave thought that much training could be done with lower fidelity devices. He pointed to successes in training during the 1950s with much less simulator capability and fidelity, and to their whole training program, which includes all media. He commented, though, that the airlines are driven by requirements of the regulatory agencies, and by the legal implications of what they do. Having achieved zero-time (flight time) training and its cost benefits, it is unlikely that any airline would change anything that might jeopardize the benefits of the whole approach. This includes using the maximum state of the art in flight simulators.

Validity of Behavior in Simulators

We asked another point-blank question: Is the behavior you have observed of pilots in a simulator any different from what you would expect in the real world? Dave seemed surprised that anyone would ask this question. For him, there is no question that the behavior in their simulators is valid. He cited an example from the days before modern simulators, visual, and motion systems, where there was an intentional gear-up landing in Los Angeles caused by a system failure. The pilot did everything perfectly, and commented that it was "a piece of cake" because he had just practiced that problem in the simulator.

This view probably is common in a commercial airline training environment, where there is high motivation to learn and maintain proficiency. We do not know if the same motivation would operate in a research simulator setting, but certainly elements of professional pride in proficiency and previously mentioned aspects of peer pressures would be operating. Together they might create motivational levels which are equivalent to those found in commercial airline training.

LOFT Guidelines

We asked if there was anything in the LOFT guidelines report (Lauber and Foushee, 1981) that is no longer true, or is out of date. Dave said the principles are just as valid today as they were then. He knew of nothing that needed to be changed, and assured us that if there was anything controversial in that report, he would certainly be aware of it. He was most pleased that the Air Force Military Airlift Command has pursued the development of LOFT exercises under the rubric of MOS training (MOST) based on the LOFT guidelines report.

APPENDIX I

FIELD INTERVIEW McDONNELL DOUGLAS CORPORATION MILITARY MISSION-ORIENTED SIMULATION RESEARCH

Dr. William J. Cody
Lead Engineer, Life Sciences Division
McDonnell Douglas Corp., St. Louis, Missouri

Dr. Cody directed three major studies for the Air Force involving use of the McDonnell Douglas flight simulator facility. Two studies had the purpose of quantifying the effects of chemical defense (CD) stressors on pilot performance. The third was an evaluation of the mission effectiveness of an F-15 Dual Role Fighter (DRF) crew system concept for a pilot and weapons systems officer. All three studies involved air-to-ground attack scenarios.

CUSTOMER RELATIONS

The customer imposed difficult constraints on the CD studies. Only six pilots were to be tested because of cost, and the collection of baseline, i.e., normal operation, data was not included because this was in effect a test of the pilots' competence.

Discussions with the customer to translate the customer's requirements to testable propositions is an important part of the initial work. The customer almost always formulates the study question in a general or practical way. It is not always clear what the research questions should be to answer the practical question. The researcher should be careful not to undertake the project without a clear understanding with the customer of what will be done in specific terms.

SCENARIO DEVELOPMENT

The main issue in the chemical defense study was how body heating equivalent to wearing individual protective equipment (IPE) would affect attack-mission performance. The heat load was imposed by an undergarment with tubing woven in for circulation of hot water. It was difficult to know how long the segments should be. The customer provided data on what the expected body heat change should be with time wearing the IPE, but it was not possible to determine experimentally ahead of time what temperature profile for heating the suit would produce the desired body temperature, or how long the mission segments should be.

In the first CD study the mission profiles originally conceived turned out in preliminary testing to be far too difficult. They had to be made less demanding to be practical. (Notice that this problem is common to the predisposition to overcomplicate LOFT or LOS scenarios.) Six different profiles were developed for the first CD study. These profiles were considered initially to be equally difficult based on the total distance flown, the number of waypoints and the number of heading changes. It turned out, however, that pilots found them to be very different in difficulty because of differing demands associated with angles to the target, distance and the weapon used. Cody said that his lack of familiarity with the details of the attack mission led him to initially oversimplify the equal-difficulty problem.

In the second CD study a different approach was taken to develop equivalent profiles. A single mission that had several legs, waypoints and two target locations was designed. A template of this profile was drawn and other profiles were generated by rotating the template with respect to the simulated terrain. Thus the specific headings, terrain path and target locations all changed, but the fundamental profile did not. Since the pilots think and perform in terms of the specific headings and ground references, they did not realize that the different missions were fundamentally the same. This was a successful means for creating differences in the appearances of the scenario, while maintaining the similarities necessary for data analysis. A similar technique was used in the DRF study.

In the DRF study, relatively short mission segments were used. The scenario began approaching the forward edge of the battle area (FEBA) and ended on the return crossing. Cody found that his experiences performing the CD studies were a great help in performing the DRF study. Length and character of the scenario were dictated by a consideration of the number of observations, i.e., data collection segments necessary and the crew functions to be performed. Tasks were segmented and approximate performance times were associated with each task segment. Laying out the segments on a time line and taking data collection needs into account produced an estimate of the length and composition of the scenario.

The simulation for the CD and DRF study were not full mission in the sense of including every aspect of the mission.

Compromises were made in the interest of economy of effort and experimental control. For instance, the mission planning phase, which can require 2-3 hr for a 1-hr mission, was not included. Also, communications were restricted because it was not clear how they could be controlled and manipulated in a way advantageous to the study. They would be a confounding factor only, and not essential to the flying of the mission. Only the systems essential to the mission were operational. This saved training time and removed additional sources of variability. Also, cockpit checklists were abbreviated to include only those systems that were essential to performing the mission. These were all acceptable reductions in mission fidelity, because the purpose of the studies was to determine how well the pilots could perform specific tasks with particular equipment and procedures.

For the CD studies, four different models were used to predict what behavioral functions related to mission performance would likely be sensitive to heat stress. The predictions from these models were used as a basis for the scenario design, i.e., to emphasize aspects of the mission that are most likely to show effects. For example, tracking performance was one of the functions predicted to be sensitive to thermal stress. Consequently, the scenario was designed to include tracking segments with wind-gust disturbances. Pilots thought these relatively long straight segments were unrealistic, but realism was compromised for the sake of obtaining useful data.

F-15 pilots were used in the preliminary testing of the DRF missions. Several changes to the scenario were made based on their comments. This reinforces the point that preliminary testing is a critical part of any simulation-based research.

To keep the scenarios simple, Cody had originally planned to use a single weapon type. The pilots objected because they said that would never be done. So to satisfy the pilots in the DRF study, appropriate weapons were paired with different targets although it was not a factor of interest in the research.

SIMULATOR ISSUES

The simulator facility is essentially modular in hardware and software. Developing the simulation thus involved linking the hardware units necessary, e.g., the cockpit and terrain board, and then assembling the proper software elements. Problems in the development of the simulation configuration included taking real-time computational demands into account. The number and type of events, as well as the data logging needs, had to be assessed to ensure the computer would not be overloaded. The parameter definitions and formatting also had to be done with care. Great detail of specification was necessary to communicate to the programmers exactly what the investigator's intentions were.

In both studies a constraint was imposed by the simulator visual system 60° field of view and the amount of terrain available on the terrain board. It took quite a bit of effort to develop a profile that would keep the target in the field of view long enough to perform the attack run, and still keep the visually guided flight portions of the mission within the limits of the terrain board.

Shortcomings of the simulation for the purpose of the studies included the inability to program event-versus-time-contingent occurrences, the relative inflexibility of the code (it takes a great deal of programming effort to change a profile), and the lack of real-time data reduction to provide summaries of a run shortly after they occurred. It was vital to keep a test director's log to record the time of start and ending of segments of interest for data analysis, and to note when some problem or other produced bad or contaminated data.

SUBJECT ISSUES

Selection and training

Three types of aircrew members were used in the DRF study. Three crews had experience in the two-seat F-15B2 aircraft. The other three crews were composed of three F-15 (single seat) pilots and three F-111 weapons systems officers. The experience of three different types of aircrew personnel affected the comprehensiveness of the simulation and the choice of tasks. A practical by-product of using the F-15B2 pilots was their exposure to the system configuration developed by the company. To the degree that their experiences in the simulator gave them an appreciation of the merit of this configuration, they would become positive advocates for it in the operational community. It, of course, made sense to use these pilots for the study because it minimized the training problem.

Issues about the interaction of subject characteristics, mission task requirements, and simulator fidelity became apparent in the CD studies. Two types of pilots were used, F-15 pilots and Air Guard A-7 pilots. An F-15 simulator was used, and the task was air-to-ground weapon delivery. F-15 pilots are familiar with the F-15 cockpit, but because this aircraft has an air-combat mission, the pilots were not proficient in ground-attack maneuvering. The A-7 pilots were unfamiliar with the F-15 cockpit, but knew ground-attack procedures very well. The trade-offs were to teach F-15 pilots to do ground attack, and familiarize the A-7 pilots with the F-15 cockpit since both types of pilots were used, the F-15 pilots in the initial study and A-7 pilots in the second. When the results of the two studies were compared, it turned out that the A-7 pilots were about twice as good in their bombing scores as the F-15 pilots.

Normally, subjects are chosen because of their familiarity with both the specific aircraft and the mission. In this case one group was familiar with the aircraft and the other with the mission. The experience here was that, in terms of mission performance, it was easier to teach the A-7 pilots to fly the F-15 than to teach the F-15 pilots to perform the mission. However, had the mission depended heavily on use of all the aircraft systems, it is likely that the F-15 pilots would have done better. The lesson is that choosing the subject population is not always a simple decision. Specific experiences must be weighed against mission requirements as well as the configuration of the simulation equipment.

In training the pilots for the attack mission, a profile that was one of the variants of the general mission profile was used. The criterion for completion of training was three successful weapons delivery runs.

Physical and Psychological Well-Being

When research involves physical stressing of the subject, the performing organization should not accept responsibility for the well-being of the subjects. In this case, the Air Force provided a medical doctor to monitor the tests and accept responsibility for the Air Force pilots. Understandably, in studies where behavioral changes or physiological effects are expected to occur, it is difficult to get informed consent from pilots.

Surface-to-air missile (SAM) threats were included in the mission profile. An issue was whether to allow the aircraft to be hit by a SAM and, if so, would the mission stop. It was essentially a realism vs. research practicality issue. If the aircraft was never hit, the pilots would soon learn that the missiles were really not a significant threat. Since these are operational pilots, there was a danger that altering their expectations of the real threat could have lethal implications. On the other hand, it would disrupt the research if the simulation terminated when the aircraft was struck by a missile.

The compromise was to allow the aircraft to be hit if the pilot did not counter the threat by ECM or maneuvering, but only cue the pilot that he was hit and not stop the mission. Unlike civilian air operations, military pilots are used to disasters, i.e., dying in mock combat. They do not like it, but do see it as a valuable training experience. Because this is routine in military training, it does not have the ethical implication of possible psychological harm to the pilot, which would be the case in simulations of civil air operations.

Subject Attitudes

F-15 pilots were reluctant to participate in the CD study because they did not want to give up aircraft flight time. The Air Guard pilots were more willing to participate because of

the opportunity to get some F-15 experience, if only in a simulator.

Experienced pilots were used in the CD and DRF studies, and are used in most of the simulation work McDonnell Douglas does. The cockpits are comprehensive and realistic representations of the actual aircraft, and the aircraft dynamics are accurate. (The F-15 and F-18 design was basically derived from simulation developments rather than the reverse.) Physical fidelity is not an issue in the simulation studies. Pilots do express some concerns about the lack of cockpit motion because the main simulators are all fixed base. The pilots had no concerns about the low luminance of the visual simulation, but were unhappy with the limited resolution and field of view. Early target detection and side viewing are important to air-to-ground attack missions and are, therefore, the likely basis of the concerns about resolution and field of view.

The main concern of the subjects is mission or scenario fidelity. Pilots do not readily accept deviations from operational practice unless the study is clearly to try new equipment and procedures. Part of the concern about scenario fidelity stems from the fact that a simulation that is essentially a duplication of an actual mission is a test of the pilots' capabilities. If the pilots are to submit to such testing, they want everything right to maximize their opportunity to properly perform. They do not want a shortcoming in the simulation to be interpreted as a lack of their ability. Conversely, the pilots are much more favorably disposed to trying new equipment and procedures because they cannot be held accountable for the outcome.

Cody commented that the pilots are very skeptical that a nonpilot psychologist can measure expert pilot behavior. It is important to include a pilot on the research team to work with the researcher and interact together with the subject pilots. The subjects then have some assurance that somebody who knows the operational world is involved, and can discuss the reasons for the characteristics of the simulation study in credible terms.

PERFORMANCE MEASUREMENT

Multiple means of performance measurement were used by Cody on the CD and DRF studies. Expert opinion of experienced pilots was used heavily during the scenario development and checkout phases. Informal dialogue between the experts and the investigator were the primary means used in these stages. The Subjective Workload Assessment Technique (SWAT) was administered routinely to the subjects since workload was a major consideration in both studies.

The subjects also were given questionnaires before and after the test sessions. The preliminary questionnaires collect biographical and experience information. The post-test questionnaires were another source of expert-opinion information

about the simulated system and procedures, as well as the simulation and scenario. The pilots also were asked to rate their own performance.

Automatic recording of system-state data was used extensively. It was the primary source of objective data, and was relied on for final analyses as much as possible. The general philosophy was to collect as much of this type of data as possible; it may turn out to be useful later. Video and audio tapes were valuable sources of information. Crew actions, communications, cockpit displays and mission track were all recorded. These were later edited together in a time-linked, split-screen format to simplify interpretation. These tapes were used for link-analysis, classification of crew communications and observer scoring of performance.

SCHEDULING

Cody noted that scheduling of simulator time was a problem and they usually worked off-hours. An informal priority of simulator use is followed. First priority is marketing, second is engineering development of specific aircraft systems, third is training company pilots, and fourth is contracted research.

The initial schedule for data collection was too tight. He tried running three pilots or crews per week, but simulator failures and subjects not showing up on time created severe, sometimes impossible, problems for completing the planned runs. Cody said that during the period of actual data collection, the investigator should plan on having about 50% of the scheduled simulator use time being used for the runs. That is, if the simulator is scheduled for 8 hr/day, no more than 4 hr/day of actual running time should be planned.

**ORIGINAL PAGE IS
OF POOR QUALITY**

BIBLIOGRAPHY

- Adams, J. A.: On the Evaluation of Training Devices. *Human Factors*, vol. 21, no. 6, 1979, pp. 711-720.
- Advisory Group for Aerospace Research and Development. *Fidelity of Simulation for Pilot Training*. AGARD-AR-159, 1980.
- Alluisi, E. A.: Methodology in the Use of Synthetic Tasks to Assess Complex Performance. *Human Factors*, vol. 9, no. 4, 1967, pp. 375-384.
- Anderson, J. R.: *Cognitive Skills and their Acquisition*. Lawrence Erlbaum Associates, 1981.
- Baer, D. K.; Jones, D. W.; and Francis, C. C.: *Nuclear Power Plant Simulators: Their Use in Operator Training and Requalification*. NRC Fin No. B0421-8, Department of Energy, 1980.
- Baron, S.; Pew, R.; and Zacharias, G.: *Performance Measurement for Simulator Evaluation*, Final Report. NASA CR-166037, 1980.
- Bartley, S. H.: What Do We Call Fatigue? *Psychological Aspects and Physiological Correlates of Work and Fatigue*, E. Simonson and P. Weiser, eds. 1976, pp. 409-417.
- Baum, D. R.: A Framework and Topics for Empirical Research on Training Simulator Effectiveness. *Research Issues in the Determination of Simulator Fidelity: Proceedings of the ARI Sponsored Workshop (Technical Report 547)*, R. T. Hays, ed. U.S. Army Research Institute for the Behavior and Social Sciences, 1981, pp. 25-38.
- Belsterling, C. A.; Pandey, S.; Felkins, L.; Praskievicz, J. C.; and Kozinsky, E. J.: *Specification and Verification of Nuclear Power Plant Training Simulator Response Characteristics: Part I. Summary of Current Practices for Nuclear and Non-Nuclear Simulators*. ORNL/TM-7985/P1. Oak Ridge National Laboratory, 1982.
- Blaiwes, A. S.; Puig, J. A.; and Regan, J. J.: *Transfer of Training and the Measurement of Training Effectiveness*. *Human Factors*, vol. 15, no. 6, 1973, pp. 523-533.
- Blake, R. R.; and Mouton, J. S.: *The New Managerial Grid*. Gulf, 1978.
- Brissenden, R. F.: *Using Simulation for Research--and Using Research to Develop Valid Simulation Techniques*. *Proceedings of the SAE 3rd International Simulation and Training Conference*, 1967, pp. 103-115.
- Carnap, R.: *Logical Foundations of the Unity of Science*. *International Encyclopedia of Unified Science*, vol. I, part 1. O. Neurath, R. Carnap, and C. W. Morris, eds. University of Chicago Press, 1955, pp. 42-62.
- Caro, P.: *Aircraft Simulators and Pilot Training*. *Human Factors*, vol. 15, no. 6, 1973, pp. 502-509.
- Caro, P. W.: *Some Factors Influencing Transfer of Simulator Training*. HumRRO-PP-1-76, Human Resources Research Organization, August 1976.
- Caro, P. W.: *Some Current Problems in Simulator Design, Testing, and Use*. HumRRO-PP-2-77. Human Resources Research Organization, March 1977.
- Clement, W. F.; Allen, R. W.; Heffley, R. K.; Jewell, W. F.; Jcx, H. R.; McRuer, D. T.; Schulman, T. M.; and Stapleford, R. L.: *Functional Requirements for the Man-Vehicle Systems Research Facility*. NASA CR-166315, 1980.
- Cody, W. J.: *Recommendations for Conducting Manned Flight Simulation Research*. McDonnell Douglas Astronautics Company, 1984.
- Cooper, G. E.; White, M. D.; and Lauber, J. K.: *Resource Management on the Flight Deck*. NASA CP-2120, 1979.
- Crawford, A.; and Brock, J. F.: *Using Simulators for Performance Measurement. Productivity Enhancement: Personnel Performance Assessment in Navy Systems*. L. T. Pope and D. Meister, eds. Navy Personnel Research and Development Center, 1977, pp. 179-191.
- Crowe, W.; Hicklin, M.; Kelly, M.; Obermayer, R.; and Satzer, W.: *Team Training Through Communications Control*. NAVTRAEQUIPC 80-C-0095-1. Navy Training Equipment Center, 1981.
- Davies, J.: *A Simulator to Train Process Operators on a Substitute for Natural Gas Plants*. *Proceedings of Human Operators and Simulation*. University of Technology, Institute of Measurement and Control, U.K., 1977, pp. 145-150.
- Eddowes, E. E.: *Proceedings of the Simulator Effectiveness Research Planning Meeting*. AFHRL-TR-77-72. Air Force Human Resources Laboratory, 1977.
- Ellis, K.; and Claridge, A. N.: *An Analytical and Experimental Study of the Design of Machinery Control Panels*. *Proceedings of Human Operators and Simulation*. University of Technology, Institute of Measurement and Control, U.K., 1977, pp. 211-220.
- Emery, J. H.; Sonneborn, W. G. O.; and Elam, C. B.: *A Study of the Validity of Ground-Based Simulation Techniques for the UH-1B Helicopter*. USAAVLABS-TR-67-72. U.S. Army Aviation Materiel Laboratories, 1967.
- Erwin, D. E., ed.: *Psychological Fidelity in Simulated Work Environments*. *Proceedings of a Symposium at the Annual Meeting of the American Psychological Association*, ARI-RES PROBLEM REV-78-26. U.S. Army Research Institute for the Behavioral and Social Sciences, 1978.
- Forbus, K.; and Stevens, A.: *Project Steamer: III. Using Qualitative Simulation to Generate Explanations of How to Operate Complex Physical Devices*. Bolt Beranek and Newman Inc., 1981.
- Geisler, M. A.; and Ginsberg, A. S.: *Man-Machine Simulation Experience*. P-3214, IBM Scientific Computing Symposium: *Simulation Models and Gaming*, Dec. 1964.

- Govindaraj, T.; and Su, Y.-L.: Training Simulator Design for a Marine Engine Room at Different Levels of Fidelity. Proceedings of the Human Factors Society 27th Annual Meeting. A. T. Pope and L. D. Haugh, eds. Santa Monica, CA., Human Factors Society, 1983, pp. 1001-1005.
- Hall, E. R.; Parker, J. F. Jr.; and Meyer, D. E.: A Study of Air Force Flight Simulator Programs. AMRL-TR-67-111, Aerospace Medical Research Laboratories, Wright-Patterson AFB, OH, 1967.
- Hammell, T. J.; and Puglisi, J. J.: Empirical Investigation of Simulator/Training System Characteristics. vol. 27, no. 2, 1980, pp. 106-115.
- Hammerton, M.: Transfer and Simulation. Proceedings of Human Operators and Simulation. University of Technology, Institute of Measurements and Control, U.K., 1977, pp. 1-6.
- Hays, R. T.: Simulator Fidelity: A Concept Paper. AD-A101475 ARI-TR-490, U.S. Army Research Institute for the Behavioral and Social Sciences, 1980.
- Hays, R. T., ed.: Research Issues in the Determination of Simulator Fidelity: Proceedings of the ARI Sponsored Workshop. ARI-TR-547, U.S. Army Research Institute for the Behavioral and Social Sciences, 1981.
- Hennessy, R. T.; Sullivan, D. J.; and Cooles, H. D.: Critical Research Issues and Visual System Requirements for a V/STOL Training Research Simulator NAVTRAEQUIPC 78-C-0076-1. Naval Training Equipment Center, 1980.
- Hennessy, R. T.; Lintern, G.; and Collyer, S. C.: Unconventional Visual Displays for Flight Training. NAVTRAEQUIPC TR-81-014, Naval Training Equipment Center, 1981.
- Hoffman, L. G.; and Riedel, S. A.: Manned Engineering Flight Simulation Validation. Part I: Simulation Requirements and Simulator Motion System Performance. AFFDL-TR-78-192, Part I. Air Force Flight Dynamics Laboratory, 1979.
- Houtman, G. M.; Laughery, K. R., Jr.; and Ditzian, J. L.: Development of a Simulator Certification (SIMCERT) Methodology for SAC. Proceedings of the Human Factors Society, 25th Annual Meeting, 1981, pp. 69-73.
- Huff, E. M.; and Nagel, D. C.: Psychological Aspects of Aeronautical Flight Simulation. *American Psychologist*, no. 3, 1975, pp. 426-439.
- Hunt, R. M.; and Rouse, W. B.: Problem-Solving Skills of Maintenance Trainees in Diagnosing Faults in Simulated Powerplants. *Human Factors*, vol. 23, no. 3, 1981, pp. 317-328.
- Hunter, S.; Gundry, A. J.; and Rolfe, J. M.: Human Factors Topics in Flight Simulation: An Annotated Bibliography. AGARD Report No. 656, 1976.
- Jones, E. R.; Hennessy, R. T.; and Deutsch, S., eds.: Human Factors Aspects of Simulation. National Academy Press, 1985.
- Keegan, J. B.: The Use of Simulators for Vehicles and Complex Systems. Proceedings of Human Operators and Simulation. University of Technology, Institute of Measurement and Control, U.K., 1977, pp. 7-18.
- Kraft, C. L.; and Elworth, C. L.: Flight Deck Workload and Night Visual Approach Performance. Document No. DG-22996. The Boeing Company, 1969.
- Krahenbuhl, G. S.; Marett, J. R.; Reuther, L. C.; Constable, S. H.; and Reid, G. B.: Pilot Stress in A-10 Surface-Attack Training. AFHRL-TR-80-16, Air Force Human Resources Laboratory, Operations Training Division, Williams AFB, AZ, 1980.
- Kuhn, T. S.: The Structure of Scientific Revolutions (2nd edition), University of Chicago Press, 1970.
- Lager, Carl: Pilot Reliability. The Royal Institute of Technology, Sweden, 1974.
- Lauber, J. K.: Cockpit Resource Management in New Technology Aircraft. National Aeronautics and Space Administration, Ames Research Center, 1981.
- Lauber, J. K.; and Foushee, H. C.: Guidelines for Line-Oriented Flight Training. Proceedings of a NASA/Industry Workshop: Volume I, NASA CP-2184, 1981.
- Lauber, J. K.; and Foushee, H. C.: Guidelines for Line-Oriented Flight Training. Proceedings of a NASA/Industry Workshop: Volume II. NASA CP-2184, 1981.
- Lew, D. W.; and Dyda, K. J.: Ground-Based Simulation Techniques. USAAVLABS-TR-67-56. U.S. Army Aviation Materiel Laboratories, 1967.
- Lintern, G.: Transfer of Landing Skills after Training with Supplementary Visual Cues. *Human Factors*, vol. 22, no. 1, 1980, pp. 81-88.
- Man-Vehicle Systems Research Facility: Design and Operating Characteristics. NASA TM 84372, 1983.
- Marshall, E.; and Shepherd, A.: Strategies Adopted by Operators When Diagnosing Plant Failures from a Simulated Control Panel. Proceedings of Human Operators and Simulation. University of Technology, Institute of Measurement and Control, U.K., 1977, pp. 59-65.
- MARSIM 81, Second International Conference on Marine Simulation. National Maritime Research Center, 1981.
- Matheny, W. G.: Investigation of the Performance Equivalence Method for Determining Training Simulator and Training Methods Requirements. AIAA Paper 75-108, 1975.
- Meeker, R. J.; and Shure, G. H.: Updating Some Ground Rules for Man-Machine Simulation. Technical Report SDC-SP-3143. System Development Corporation, 1968.
- Meister, D.: Behavioral Analysis and Measurement Methods. John Wiley and Sons, 1985.
- Montague, W. E.: Is Simulation Fidelity the Question? Research Issues in the Determination of Simulator Fidelity: Proceedings of the ARI Sponsored Workshop, R. T. Hays, ed. TR-547, U.S. Army Research Institute for the Behavioral and Social Sciences, 1981, pp. 93-109.

- Morris, C. W.: Scientific Empiricism. International Encyclopedia of Unified Science, vol. I, part 1, O. Nerath, R. Carnap, and C. W. Morris, eds. University of Chicago Press, 1955, pp. 63-75.
- Muckler, F. A.; Nygaard, J. E.; O'Kelly, L. I.; and Williams, A. C., Jr.: Psychological Variables in the Design of Flight Simulators. WADC TR 56-369, Wright Air Development Center, 1959.
- Murphy, Miles R.; and Awe, C. A.: Aircrew Coordination and Decision-Making: Peer Ratings of Video Tapes Made During a Full-Mission Simulation. NASA CP-2428, 1985, pp. 23.1-23.33.
- Murphy, M. R.; Randle, R. J.; Tanner, T. A.; Frankel, R. M.; Goguen, J. A.; and Linde, C.: A Full-Mission Simulator Study of Aircrew Performances: The Measurement of Crew Coordination and Decisionmaking Factors and Their Relationships to Flight Task Performances. Proceedings of the 20th Annual Conference on Manual Control, vol. II. NASA Ames Research Center, 1984, pp. 249-262.
- Nagel, D. C.; and Shiner, R. J.: The Man-Vehicle Systems Research Facility: A New NASA Aeronautical R&D Facility. AIAA Paper 83-1098, 1983.
- National Research Council: Automation in Combat Aircraft. National Academy Press, 1982.
- National Research Council: Report of the Committee to Assess the Computer-Aided Operations Research Facility, 1983.
- Obermeyer, R. W.; Slemon, G. K.; Johnston, D. L.; and Hicklin, M. B.: Team Training: Knowledge-Based Simulation for Team Members. NAVTRAEQUIPC 82-C-0140-1. Naval Training Equipment Center, 1984.
- Orlansky, J.: Research on the Fidelity of Simulators. Research Issues in the Determination of Simulator Fidelity: Proceedings of the ARI Sponsored Workshop, R. T. Hays, ed. TR-547. U.S. Army Research Institute for the Behavioral and Social Sciences, 1981, pp. 143-151.
- Owen, D. H.; and Jensen, R. S.: Methodological Approaches to Identifying Relevant Features for Visual Flight Simulation. AFOSR-TR-81-0479, Air Force Office of Scientific Research, 1981.
- Perrow, C.: The Organizational Context of Human Factors Engineering. Administrative Science Quarterly, vol. 28, no. 4, 1983, pp. 521-541.
- Pew, R. W.; Fehrer, C. E.; Baron, S.; and Miller, D. C.: Critical Review and Analysis of Performance Models Applicable to Man-Machine Systems Evaluation. AFOSR-TR-77-0520. Air Force Office of Scientific Research, 1977.
- Pope, L. T.; and Meister, D., eds.: Productivity Enhancement: Personnel Performance Assessment in Navy Systems. Naval Personnel Research and Development Center, 1977.
- Popplow, J. R.: After the Fire-Ball. Aviation, Space, and Environmental Medicine, vol. 55, no. 4, 1984, pp. 337-338.
- Prophet, W. W.; Shelnutt, J. B.; and Spears, W. D.: Simulator Training Requirements and Effectiveness Study (STRES): Future Research Plans, AFHRL-TR-80-37. Air Force Human Resources Laboratory, Wright-Patterson AFB, OH, 1981.
- Provenmire, H. K.; and Roscoe, S. N.: Incremental Transfer Effectiveness of a Ground-Based General Aviation Trainer. Human Factors, vol. 15, no. 6, 1973, pp. 534-542.
- Rasmussen, J.: Skills, Rules and Knowledge: Signals, Signs, and Symbols, and Other Distinctions in Human Performance Models. IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-13, 1983, pp. 257-266.
- Rockwell, T. H.; and Griffin, W. C.: An Investigation into Pilot and System Response to Critical In-Flight Events, Vol. I: Executive Summary. NASA CR-164993, 1981.
- Rockwell, T. H.; and Griffin, W. C.: An Investigation into Pilot and System Response to Critical In-Flight Events, Vol. II: Appendix. NASA CR-164994, 1981.
- 50 Years of Flight Simulation (Vols. I, II, and III). The Royal Aeronautical Society, 1979.
- Roscoe, S. N.: Review of Flight Training Technology. ARI-RES PROBLEM REV-76-3. U.S. Army Research Institute for the Behavioral and Social Sciences, 1976.
- Rouse, W. B.; and Rouse, S. H.: A Note on Evaluation of Complex Man-Machine Systems. IEEE Transactions on Systems, Man and Cybernetics, vol. SMC-14, no. 4, 1984, pp. 633-636.
- Ruffell-Smith, H. P.: A Simulator Study of the Interaction of Pilot Workload with Errors, Vigilance, and Decisions. NASA TM-78482, 1979.
- Ruocco, J.; Klier, S.; Gage, H.; and Vitale, P.: Design Factors in Environmental Simulation. NAVTRADEVCE 66-C-0019-1. Naval Training Device Center, 1970.
- Semple, C. A.: Simulator Training Requirements and Effectiveness Study (STRES). AFHRL-TR-80-63. Air Force Systems Command, 1981.
- Semple, C. A.; Cotton, J. C.; and Sullivan, D. J.: Aircrew Training Devices: Instructional Support Features. AFHRL-TR-80-58. Air Force Systems Command, 1981.
- Shepherd, A.: Fidelity of Simulation for Training Control Panel Diagnosis. Proceedings of Human Operators and Simulation. University of Technology, Institute of Measurement and Control, U.K., 1977, pp. 129-135.
- Sheridan, T. B.; Fischhoff, B.; Posner, M.; and Pew, R. W.: Research Needs for Human Factors. Pt. 4: Supervisory Control Systems, 1983, pp. 1-43.
- Sheridan, T. B.; and Hart, S. G.: Pilot Workload, Performance and Aircraft Control Automation. Paper for NATO-AGARD Aerospace Medical Panel Symposium, 1984.
- Sheridan, T. B.; and Hennessy, R. T., eds.: Research and Modeling of Supervisory Control Behavior. National Academy Press, 1984.
- Sheridan, T. B.; and Johannsen, G., eds.: Monitoring Behavior and Supervisory Control. Plenum, 1976.

- Shettel, H. H.; Schumacher, S. P.; and Gatewood, R. D.: Driver Training Simulators, Ranges and Modified Cars: A Review and Critique of the Experimental Literature. DOT/HS-800 644. U.S. Department of Transportation, 1971.
- Simon, C. W.; and Roscoe, S. N.: Application of a Multifactor Approach to Transfer of Training Research. NAVTRAEQUIPC 78-C-0060-6, Navy Training Equipment Center, 1981.
- Singleton, W.: The Model Supervisor Dilemma. Monitoring Behavior and Supervisory Control, T. Sheridan and G. Johanssen, eds. Plenum, 1976, pp. 261-270.
- Society for Applied Learning Technology and American Institutes for Research, Mass. Conference on Simulation and Training Technology for Nuclear Power Plant Safety, 1981.
- Spears, W.: Processes of Skill Performance: A Foundation for the Design and Use of Training Equipment. NAVTRAEQUIPC 78-C-0113-4, Naval Training Equipment Center, 1983.
- Starr, C.: Progress Towards Better Simulation of Transport Aircraft. International Journal of Aviation Safety, vol. 1, no. 4, 1984, pp. 425-435.
- Stevens, A.; and Roberts, B.: Quantitative and Qualitative Simulation in Portable Training Devices. J. Computer-Based Instruction, vol. 10, no. 1-2, 1983, pp. 16-19.
- Stevens, A.; Roberts, B.; and Stead, L.: The Use of a Sophisticated Graphics Interface in Computer-Assisted Instruction. IEEE Computer Graphics and Applications, vol. 3, no. 2, 1988, pp. 25-31.
- Sulzer, R. L.; Cox, W. J.; and Mohler, S. R.: Flight Crew Member Workload Evaluation. Cited without further details by Wainwright, W. A. in Workload Evaluation. International Journal of Aviation Safety, Mar. 1984, p. 420.
- Utall, W. R.: A Taxonomy of Visual Processes. Erlbaum, 1981.
- Valverde, H. H.: A Review of Flight Simulator Transfer of Training Studies. Human Factors, vol. 15, no. 6, 1973, pp. 510-523.
- Van Gaasbeek, J. R.: Validation of the Rotorcraft Flight Simulation Program (C81) Using Operational Loads Survey Flight Test Data. USAAVRADCOM-TR-80-D-4. U.S. Army Research and Technology Laboratories, 1980.
- Vreuls, D.; Wooldridge, A. L.; Obermayer, R. W.; Johnson, R. M.; Norman, D. A.; and Goldstein, I.: Development and Evaluation of Trainee Performance Measures in an Automated Instrument Flight Maneuvers Trainer. NAVTRAEQUIPC 74-C-0063-1. U.S. Naval Training Equipment Center, 1975.
- Vreuls, D.; and Wooldridge, L.: Aircrew Performance Measurement. Productivity Enhancement: Personnel Performance Assessment in Navy Systems. L. T. Pope and D. Meister, eds. Navy Personnel Research and Development Center, 1977, pp. 5-31.
- Vreuls, D.; Obermayer, R. W.; Wooldridge, A. L.; and Kelly, M. J.: Performance Measurement Guidelines for Research. AFOSR-85-0642TR. Vreuls Research Corporation, 1985a.
- Vreuls, D.; and Obermayer, R. W.: Human-System Performance Measurement in Training Simulators. Human Factors, vol. 27, no. 3, 1985b, pp. 241-250.
- Walden, D. A.; and Gress, R. K.: A Survey of Coast Guard Simulator Development Research and Simulator Applications. Proceedings MARSIM81, Second International Conference on Marine Simulation, vol. I. National Maritime Research Center, Computer Aided Operations Research Facility, 1981.
- Wickens, C. D.: The Multiple Resources Model of Human Performance: Implications for Display Design. AGARD-CP-371, 1984.
- Wiener, E. L.; and Curry, R. E.: Flight Deck Automation: Promises and Problems. NASA TM-81206, 1980.
- Williges, B. H.; Roscoe, S. N.; and Williges, R. C.: Synthetic Flight Training Revisited. Human Factors, vol. 15, no. 6, 1973, pp. 543-560.
- Williges, R.; and Mills, R.: Catalog of Methodological Considerations for Systems Experimentation. Air Force Aerospace Medical Laboratory, 1982.
- Wood, M. E.: The Fidelity Issue in Visual Simulation. Proceedings of the 1977 Image Conference. Air Force Systems Command, 1977, pp. 291-294.
- Woods, D.: Evaluation of Person-Machine System Performance, 1984.

Report Documentation Page

1. Report No. NASA TM-88330		2. Government Accession No.		3. Recipient's Catalog No.	
4. Title and Subtitle Using Full-Mission Simulation for Human Factors Research in Air Transport Operations				5. Report Date July 1988	
				6. Performing Organization Code	
7. Author(s) Harry W. Orlady (Orlady Associates, Inc., Los Gatos, CA), Robert T. Hennessy, Richard Obermayer, and Donald Vreuls (Vreuls Research Corporation, Thousand Oaks, CA), and Miles R. Murphy				8. Performing Organization Report No. A-86320	
				10. Work Unit No. 505-67-11	
9. Performing Organization Name and Address Ames Research Center Moffett Field, CA 94035				11. Contract or Grant No.	
				13. Type of Report and Period Covered Technical Memorandum	
12. Sponsoring Agency Name and Address National Aeronautics and Space Administration Washington, DC 20546-0001				14. Sponsoring Agency Code	
15. Supplementary Notes Point of Contact: Aerospace Human Factors Research Division, MS 239-1, Ames Research Center, Moffett Field, CA 94035, (415) 694-5726 or FTS 464-5726					
16. Abstract <p>This study examined state-of-the-art mission-oriented simulation and its use in human factors research. Guidelines were developed for doing full-mission human factors research on crew member behavior during simulated air transport operations. The existing literature was reviewed. However, interviews with experienced investigators who have addressed the myriad of issues involved in applied, mission-oriented research provided the most helpful information.</p> <p>The fundamental scientific and practical issues of behavioral research in a simulation environment are discussed. Guidelines are presented for planning, scenario development, and the execution of behavioral research using full-mission simulation in the context of air transport flight operations. Research is recommended to enhance the validity and productivity of full-mission research by: 1) validating the need for high-fidelity simulation of all major elements in the operational environment, 2) improving methods for conducting full-mission research, and 3) examining part-task research on specific problems through the use of vehicles which contain higher levels of abstraction (and lower fidelity) of the operational environment.</p>					
17. Key Words (Suggested by Author(s)) Mission-oriented simulation Full-mission simulation Flightcrew performance Human Factors Research Behavioral research			18. Distribution Statement Unclassified -- Unlimited Subject category -- 03		
19. Security Classif. (of this report) Unclassified		20. Security Classif. (of this page) Unclassified		21. No. of pages 88	
				22. Price A05	

